**University of Bath**

**PHD**

# The 2-Oxoacid Dehydrogenase Multienzyme Complex from Thermophilic Archaea

Marrott, Nia

*Award date:*
2012

*Awarding institution:*
University of Bath

Link to publication

## Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

Copyright of this thesis rests with the author. Access is subject to the above licence, if given. If no licence is specified above, original content in this thesis is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC-ND 4.0) Licence (https://creativecommons.org/licenses/by-nc-nd/4.0/). Any third-party copyright material present remains the property of its respective owner(s) and is licensed under its existing terms.

**Take down policy**
If you consider content within Bath's Research Portal to be in breach of UK law, please contact: openaccess@bath.ac.uk with the details. Your claim will be investigated and, where appropriate, the item will be removed from public view as soon as possible.

Download date: 06. May. 2023

# The 2-oxoacid dehydrogenase multienzyme complex from thermophilic Archaea

**Nia Laura Marrott**

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

September 2012

**COPYRIGHT**

Attention is drawn to the fact that copyright of this thesis rests with its author. A copy of this thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and must not copy it or use material from it except as permitted by law or with the consent of the author.

This thesis may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purposes of consultation.

# Acknowledgements

I would like to thank my supervisors Prof. Michael Danson and Dr. David Hough for the opportunity to work on this project, and their continued support throughout. I would also like to acknowledge Dr. Jean van den Elsen and Dr. Susan Crennell for giving me the opportunity to undertake all of the structural biology work. I would like to thank all of the current and past members of Lab 1.33 for their help, support and patience; particularly Dr. Karl Payne who guided me through the first year of my PhD and supervised my BSc. project. I am also grateful to the rest of the members of the Centre for Extremophile Research, who provided an outsiders view on my methods and results.

I would also like to thank all those with whom I have collaborated with; their time and expertises have been invaluable to my project. I would like to thank Prof. Steve Halford and Dr Jacqueline Marshall from Bristol University for performing all the analytical ultracentrifugation experiments; I am especially appreciative of the time Jacqueline spent analysing and explaining the results. I would also like to thank Dr. Dmitri Svergun, Dr. Peter Konarev, Dr. Gundolf Schenk at EMBL (Hamburg) for their help and expertise in performing and analysing all the small-angle X-ray scattering experiments.

Finally I would like to thank David Marrott and my family for their patience and support. They have provided the encouragement I have needed to continue working to the best of my ability and not to give up if things don't always go to plan.

## Table of Contents

# Abstract

The 2-oxoacid dehydrogenase multienzyme complexes (OADHCs) catalyse the following reaction:

$$2\text{-Oxoacid} + NAD^+ + CoASH \rightarrow Acyl\text{-}SCoA + CO_2 + NADH + H^+$$

These 5-10 MDa complexes comprise multiple copies of three enzymes: E1, E2, and E3. The structural core of the OADHC is usually composed of either 24 (cubic) or 60 (dodecahedral) E2 polypeptides, around which the E1 and E3 enzymes associate non-covalently.

Due to a lack of detectable activity in cell extracts, it was originally thought that Archaea did not possess any OADHCs; instead they utilised the smaller ferredoxin oxidoreductase (FOR) family of enzymes. However, the OADHC genes have since been identified in an increasing number of aerobic archaea, and given that the genomes of these organisms have been streamlined to remove any unnecessary genes, these two observations suggest that archaea may utilise both FORs and OADHC.

The *Thermoplasma acidophilum* OADHC component enzymes have been recombinantly expressed and assembled, resulting in a thermostable and thermoactive branched-chain OADHC. This complex provides a model to investigate stability, not only of the individual enzymes but also of a multi-protein complex and its assembly. The structure of this hyper-thermostable E2 core is crucial to the thermostability of the entire complex. The *Tp. acidophilum* E2 structure has been solved by analytical ultra-centrifugation, X-ray crystallography and small-angle X-ray scattering; these data have shown that this E2 assembles into a novel 42-mer structure, comprising 14 trimers, with both square and pentagonal faces, rather than either of the expected cubic 24-mer or dodecahedral 60-mer structures.

Analysis of the *Tp. acidophilum* E2 structure has identified an isoleucine 'anchor' residue that is key to the formation of the trimer-trimer interactions. The isoleucine side-chain extends from the C-terminal helix of one trimer into a hydrophobic pocket of the adjacent trimer and *vice versa*. The putative anchor residue has been removed by the insertion of a premature stop codon, resulting in E2 trimers that no longer assemble into a 42mer. This mutation has no affect on the tertiary structure of the trimers or its acyltransferase activity; moreover, the trimeric E2 retains the ability to bind the E1 and E3 enzymes, resulting in the first example of a functional trimeric OADHC.

# List of abbreviations

| Abbreviation | Definition |
|---|---|
| ΔH | Change in enthalpy |
| ΔS | Change in entropy |
| A | Arrhenius constant |
| $A_{280}$ | Absorbance at 280nm |
| aa | Amino acid |
| Abs | Absorbance |
| APS | Ammonium persulfate |
| AUC | Analytical ultracentrifugation |
| BSA | Bovine serum albumin |
| CoA/CoASH | Coenzyme A |
| DCPIP | 2,6-dichlorophenolindophenol |
| DE3 | Host has a chromosomal copy of the T7 RNA polymerase gene under control of the *lacUV5* promoter |
| DHlipoamide/DHlip | Dihydrolipoamide |
| dNTP | Deoxyribonucleotide |
| DTNB | 5',5'-Dithiobis-2-nitrobenzoate |
| E1 | 2-Oxoacid decarboxylase |
| E2 | Dihydrolipoamide transacetylase |
| E2-trunc | Truncated version of the whole E2, missing the final 5 C-terminal residues (IIYEI) |
| E2cat | E2 catalytic domain |
| E2cat-trunc | Truncated version of the E2 catalytic domain, missing the final 5 C-terminal residues (IIYEI) |
| E3 | Dihydrolipoamide dehydrogenase |
| Ea | Activation energy |
| EDTA | Ethylenediaminetetraacetic acid |
| G° | Standard Gibbs free energy |
| h | Hour |
| His-tag | Polyhistidine-tag |
| HRP | Horseradish peroxidise |
| IgG | Immunoglobulin G |
| IPTG, | Isopropyl-β-D-thio-galactoside |
| Kb | Kilobase of DNA |
| $K_B$ | Boltzmann constant |
| *k*cat | Catalytic constant |
| $K_m$ | Michaelis constant |
| LB | Luria Bertani |
| Ln | Natural logarithm |
| LplA-CTD | Lipoate-protein ligase A and C-terminal domain |
| min | Minute |
| $M_w$ | Molecular weight |
| NAD | Nicotinamide adenine dinucleotide |
| OADHC | 2-Oxoacid dehydrogenase complex |

| | |
|---|---|
| OGDHC | 2-Oxoglutarate dehydrogenase complex |
| OD | Optical density |
| PBS | Phosphate buffered saline |
| PCR | Polymerase chain reaction |
| PDB | Protein Data Bank |
| PDHC | Pyruvate dehydrogenase complex |
| PEI | Polyethylenimine |
| PSBD | Peripheral subunit binding domain |
| PTA | Phosphotransacetylase |
| R | Universal gas constant |
| r.p.m. | Revolutions per minute |
| SAXS | Small-angle X-ray scattering |
| SDS | Sodium dodecyl sulphate |
| SDS-PAGE | Sodium dodecyl sulphate polyacrylamide gel electrophoresis |
| s | Second |
| TEMED | Tetramethylethylenediamine |
| $T_{opt}$ | Temperature optimum |
| TPP | Thiamine pyrophosphate |
| U | Enzyme units - 1U is defined as the amount of the enzyme required to convert 1µmole of substrate to product per min. |
| $V_{max}$ | Maximum velocity |
| YNB medium | Yeast nitrogen base medium |

# Chapter 1: General introduction

## 1.1 Archaea

### 1.1.1 The three Domains of life

The three Domains of Life comprise Bacteria, Archaea and Eukarya. This classification system was proposed by Carl Woese et al. (1990) based on 16S ribosomal RNA sequences. The three Domains classification system replaces the phylogenetically inaccurate five Kingdom system, in which Archaea and Bacteria were grouped together, as Prokaryotes, owing to their similar appearances. Molecular characterisation of the Archaea supports their being placed on their own distinct evolutionary branch, separate from Bacteria and closer to the Eukarya branch (Figure 1.1) (Winker and Woese, 1991).

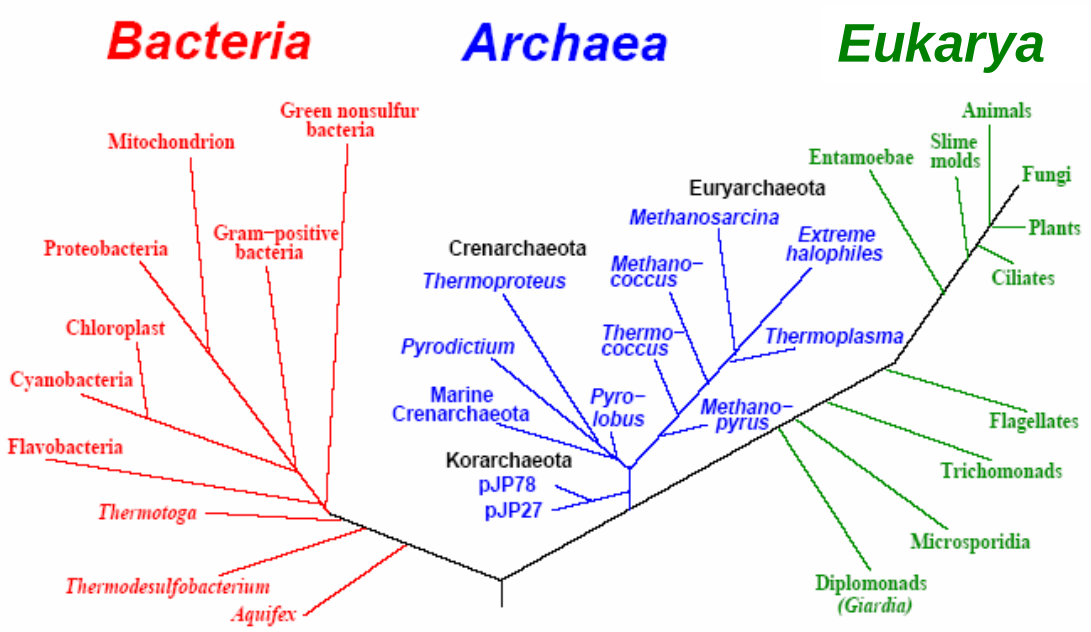



**Figure 1.1 Phylogenetic tree, based on the 16S rRNA sequences, illustrating the three Domains of life: Bacteria, Archaea and Eukarya** (Jurgens, 2002).

The Domain Archaea comprise two main phyla: Crenarchaeota and Euryarchaeota; a further two phyla, Korarchaetota and Nanoarchaeota, have tentatively been proposed based on unculturable environmental samples (Barns et al., 1996, Huber et al., 2002). Metagenomics is now beginning to identify a vast number of unculturable archaeal species; it has been suggested that up to 20% of all picoplankton cells in the world's oceans are Crenarchaeota (Karner et al., 2001). The exact arrangement of the phyla within the archaea may become more accurately described as an increasing number of archaeal species are identified.

### 1.1.2 Enzymes from extremophiles

Species of archaea and bacteria can be found in some of the world's most hostile environments, including extreme temperatures (-2°C to 15°C and 60°C to 110°C), high salinity (2-5M NaCl) and extreme pH (<4 and >9) (Hough and Danson, 1999). These extremophiles have evolved to thrive in these conditions; as a result they are now a source of novel extremophilic enzymes (extremozymes), of interest to both industry and academia. Enzymes from extreme temperatures have been characterised in an attempt to understand protein thermostability, and exploited due to their industrially favourable properties. Extremozymes currently in regular use include: the thermophilic *Taq* DNA polymerase from *Thermus aquaticus,* and the psychrophilic Cpn10 and Cpn60 chaperonins from *Oleispira antarctica* that are utilised by the ArcticExpress™ *E. coli* expression system (Stratagene. U.K.) in order to increase the recombinant protein yield. The production method of the industrial thermophilic enzymes can vary; for example, the starch-processing industry utilises recombinant amylases and other hydrolases in the conversion of starch into high-fructose corn syrup (Vieille and Zeikus, 2001); whereas, the thermophilic bacterium *Geobacillus thermoglucosidasius* is used as a whole organism, by TMO Renewables Ltd, to generate bioethanol from waste carbohydrates (Cripps et al., 2009).

#### *1.1.2.1 Protein folding, unfolding and stability*

The process of protein thermal denaturation can be described as a reversible equilibrium between the folded (F), and unfolded (U) form, coupled to the essentially irreversible conversion of the unfolded protein to the thermally denatured (D) form:

$$F \leftrightarrows U \rightarrow D$$

Figure 1.2 illustrates the activation energy (ΔG*) of the inter-conversion between the folded and unfolded forms of a protein; that is, ΔG* determines the rate at which the equilibrium is reached and therefore influences the kinetic stability of a protein. The free energy difference (ΔG) between the folded and unfolded forms of a protein determines the position of the equilibrium, and thus describes the thermodynamic stability of a protein. The ΔG value is influenced by the entropy, enthalpy and temperature of the unfolding process, shown by the following equation:

$$\Delta G_{(unfold)} = \Delta H_{(unfold)} - T\Delta S_{(unfold)}.$$

Protein thermostability can be increased by:

1) Increasing the free energy of the unfolded form, this makes U less thermodynamically favourable compared to F.

2) Increasing the ΔG*, this decreases the rate of conversion between F and U, resulting in a more kinetically stable protein.

3) Decreasing the free energy of the folded form, this increases the ΔG and ΔG* making the protein more thermodynamically and kinetically stable (Figure 1.2) (Daniel et al., 2008).

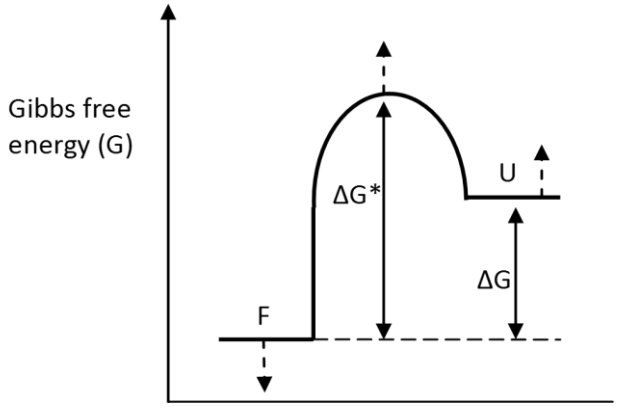



**Figure 1.2 Gibbs free energy diagram illustrating the thermodynamic and kinetic contributions to protein stability.** F and U denote the folded and unfolded protein respectively. The free energy change (ΔG) and activation energy (ΔG*) of this equilibrium are shown by the black arrows. The dashed arrows indicate energy changes required to increase the stability of a protein.

### *1.1.2.2 Structural basis of protein thermostability*

The structural basis of protein thermostability has been investigated by comparisons of homologous proteins from psychrophilic, mesophilic and thermophilic organisms. The three dimensional structures of these different temperature homologs are often similar, leading to the suggestion that thermostability is achieved by a number of small subtle changes, rather than larger global changes in the proteins structure, (Bell et al., 2002, Daniel et al., 2008) including:

- Increased hydrophobicity,
- Increased number of ionic bonds,
- Increased compactness,
- Reduced loop length and flexibility,
- Stronger N- and C-termini interactions,
- Reduced usage of thermolabile amino acids

Protein hydrophobicity

Thermostable proteins tend to have an increased number of buried hydrophobic residues (Vieille and Zeikus, 2001); the increased hydrophobicity stabilises the structure both enthalpically, through the formation of van der Waals bonds, and entropically, as the bulk water becomes increasingly disordered as a result of the burial of hydrophobic side chains. Increased hydrophobicity is less significant in the stabilization of hyper-thermostable proteins; above 80°C the bulk water becomes more disordered, and therefore the entropy increase as a result of the burial of the hydrophobic residues becomes less significant.

Ionic interactions

Ionic interactions and networks, unlike hydrophobic interactions, increase in strength at higher temperatures. The formation of a single ionic interaction is a balance between the energy required to desolvate two ions, the gain in stability as a result of the formation of a salt bridge, and any changes to the protein region surrounding the ion:

$$\Delta G_{Total} = \Delta G_{Desolvation} + \Delta G_{Electrostatic} + \Delta G_{protein}$$

The formation of a single bond requires the desolvation of two ions, and thus at lower temperatures a single ionic interaction may have little effect on protein stability. At higher temperatures, ionic interactions become increasingly important to protein stabilisation; as the surrounding water becomes more disordered the desolvation penalty decreases, and therefore the formation of an ionic bond becomes more favourable. The formation of ionic networks is another method of decreasing the desolvation penalty; ionic networks only require the desolvation of a single additional ion to create a second ionic bond.

Packing density and conformational flexibility

Thermophilic enzymes tend to be more compact than their mesophilic homologs; this increases the stability of the folded form of the protein, by increasing the number of van der Waals interactions within the core of the protein (Vieille and Zeikus, 2001). Flexible surface loops or free N- and C- termini can be initiator sites for protein thermal denaturation. Thermophilic proteins tend to have shorter loops and termini than their mesophilic homologs; loop flexibility can also be reduced by extra interactions or the presence of additional metal ions not found in mesophilic homologs (Vieille and Zeikus, 2001, Li et al., 2005). The reduced conformational flexibility of thermophilic proteins at

lower temperature enables them to remain stable at higher temperatures, preserving catalytic activity and specificity (Daniel et al., 2008).

Amino acid bias

Amino acid usage varies with temperature; above 85°C the amino acids cysteine, methionine, serine, threonine, asparagine, and glutamine can be chemically degraded if they are free in solution. However, these residues are still used by hyper-thermophilic proteins, but they tend to be found in conformationally-stable locations and shielded from the surrounding solvent (Daniel et al., 2008). An increase in the number of proline residues can stabilise a protein; prolines can only adopt a limited number of conformations, this limits the entropy of the unfolded protein, increasing the favourability of the folded form (Vieille and Zeikus, 2001). Increasing the number of intra-polypeptide disulphide bonds is another technique used to decrease the flexibility of the unfolded protein and thus promote refolding (Li et al., 2005).

### 1.1.3 *Thermoplasma acidophilum*

*Thermoplasma acidophilum* is a thermophilic Euryarchaeota, first isolated in 1970 from a coal refuse pile at the Friar Tuck mine in south-western Indiana, by Darland et al. (1970). *Tp. acidophilum* is an aerobic archaea that grows optimally at 60°C and between pH 1-2; however, it should be noted that the cell's internal pH is approximately pH 5.5 (Searcy, 1976). The 1.5Mb genome of *Tp. acidophilum* has been sequenced, and 1509 open reading frames identified (Ruepp et al., 2000). The *Tp. acidophilum* genes encoding a predicted 2-oxoacid dehydrogenase multienzyme complex have been identified, cloned and recombinantly expressed; details of this work are discussed in Chapter 1.3.

## 1.2 Carbon metabolism

The pathways involved in central carbon metabolism have many common features in all three-domains of life; these pathways not only catabolise compounds to release energy, but are also anabolic in nature, creating many of the starting substrates required for biosynthesis. This project focuses on the 2-oxoacid dehydrogenase multienzyme complex (OADHC) family of enzymes that are involved in central carbon metabolism.

There are several enzyme pathways that convert glucose into pyruvate, the three main ones are: the classical Embden-Meyerhof (EM), the non-phosphorylative Entner-Doudoroff (ED) or the pentose P pathway (Danson et al., 2007). The resulting pyruvate is then converted into acetyl-CoA by the action of either a pyruvate dehydrogenase multienzyme complex or a pyruvate ferredoxin oxidoreductase (FOR) before it enters the citric acid cycle. The citric acid cycle is a series of oxidative reactions carried out by aerobic organisms to generate $NADH+H^+$ and $FADH_2$, coupled with the release of $CO_2$ (Figure 1.3). The re-oxidation of the $NADH+H^+$ and $FADH_2$ via the electron transport chain then allows the efficient generation of ATP.

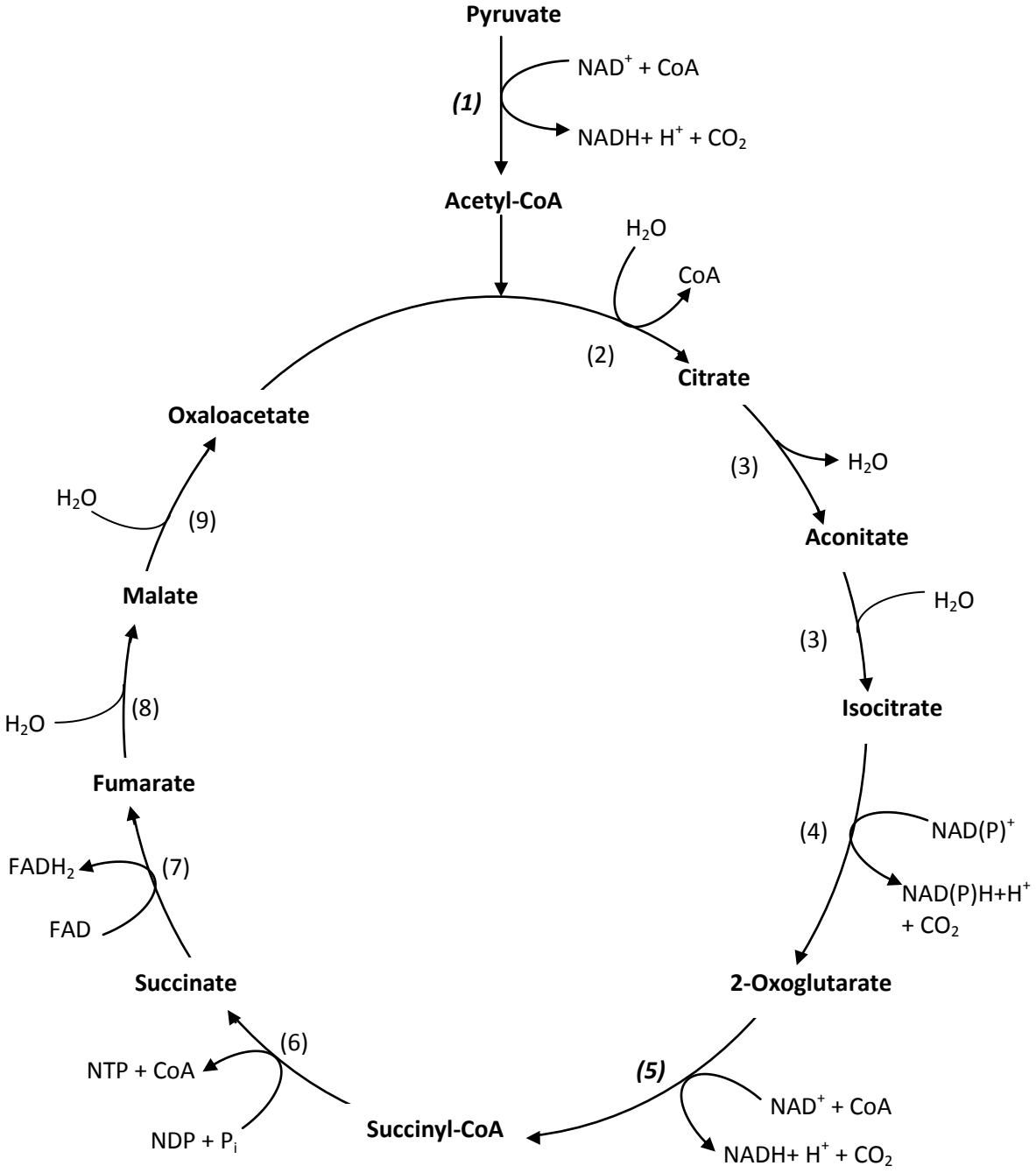



**Figure 1.3. The citric acid cycle.** Enzymes that catalyse each step are: ***[1]*** Pyruvate dehydrogenase complex/FOR, [2] Citrate synthase, [3] Aconitase, [4] Isocitrate dehydrogenase, ***[5]*** 2-Oxoglutarate dehydrogenase complex/FOR, [6] Succinate thiokinase, [7] Succinate dehydrogenase, [8] Fumarase, [9] Malate dehydrogenase. The numbers in bold italics are the reactions catalysed by either a 2-oxoacid dehydrogenase multienzyme complex or a ferredoxin oxidoreductase.

### 1.2.1 The conversion of 2-oxoacids to acyl-CoAs

Within central metabolism there are several steps that require the conversion of a 2-oxoacid to its corresponding acyl-CoA; these reactions can be catalysed by two different enzymatic systems, either the 2-oxoacid ferredoxin oxidoreductase or the 2-oxoacid dehydrogenase multienzyme complex. The structure and reaction mechanism of these two enzymatic systems are discussed later in Sections 1.2.3 and 1.2.4, respectively.

There are four main reactions requiring the conversion of a 2-oxoacid into its corresponding acyl-CoA (Figure 1.4). The first links glycolysis with the citric acid cycle, with pyruvate being converted into acetyl-CoA. The second reaction is found within the citric acid cycle, and involves the conversion of 2-oxoglutarate into succinyl-CoA. A third OADHC produces acetyl-CoA from acetoin; acetoin is a neutral molecule that can be used to store pyruvate and prevent over acidification of the cytoplasm by balancing the ratio of $NAD^+$ and $NADH+H^+$. The branched-chain amino acids, valine, leucine and isoleucine, can be deaminated to produce branched-chain 2-oxoacids, which can then be converted into their corresponding acyl-CoAs and on via a series of reactions into acetyl-CoA, acetoacetate or succinyl-CoA.

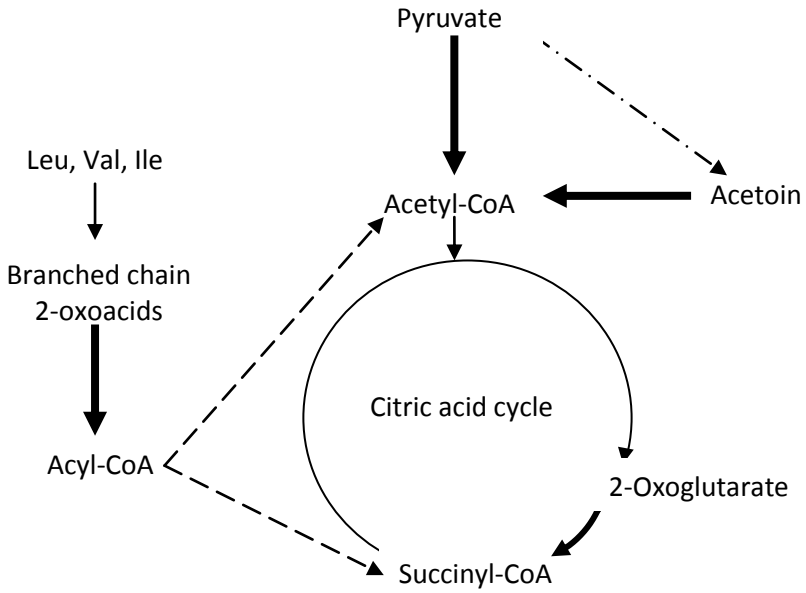



**Figure 1.4 Diagram of central metabolism indicating the location of the conversions of 2-oxoacids to their corresponding acyl-CoAs.** The reactions highlighted by bold black arrows are catalysed by either a ferredoxin oxidoreductase or 2-oxoacid dehydrogenase multienzyme complex highlighted. The dashes lines indicate the multiple reactions required to convert the branched chain acyl-CoAs into molecules such as acetyl-CoA and succinyl-CoA that can then enter the citric acid cycle. The dashed and dotted line indicates the multiple reactions required to convert pyruvate to acetoin.

### 1.2.2 The branched-chain amino acids and their corresponding 2-oxoacids and acyl-CoAs

The branched-chain 2-oxoacids are produced by the deamination of the branched-chain amino acids Val, Leu and Ile; these 2-oxoacids are then converted into their corresponding acyl-CoAs, the chemical structures of each of these molecules are shown in Table 1.1. The branched-chain acyl-CoAs then enter into the citric acid cycle via a specific pathway for each acyl-CoA; these pathways are shown in Figure 1.4. Leucine and isoleucine can be converted into acetyl-CoA and either acetoacetate or succinyl-CoA, and valine is converted into succinyl-CoA. The resulting succinyl-CoA and acetyl-CoA can then enter the citric acid cycle.

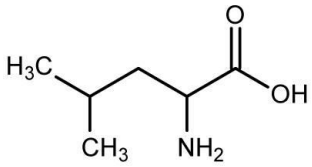
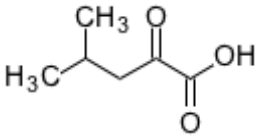
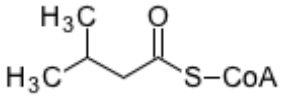
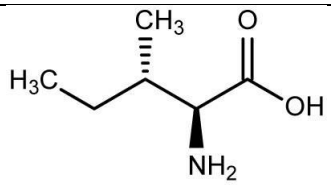
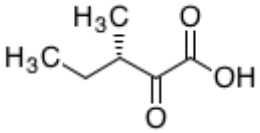
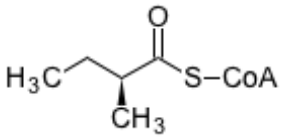
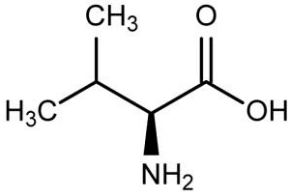
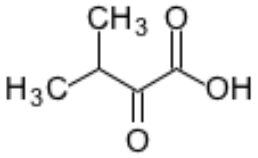
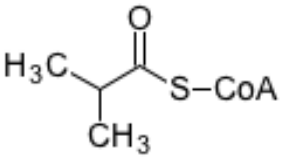
| **Amino acid** | **Branched-chain 2-oxoacid** | **Acyl-CoA** |
|---|---|---|
| Leucine | 4-Methyl-2-oxopentanoic acid | 3-Methylbutanoyl-CoA |
|  |  |  |
| Isoleucine | 3-Methyl-2-oxopentanoic acid | 2-Methylbutanoyl-CoA |
|  |  |  |
| Valine | 3-Methyl-2-oxobutanoic acid | 2-Methylpropanoyl-CoA |
|  |  |  |

**Table 1.1 Summary of the structure of the branched-chain amino acid, and the corresponding 2-oxoacid and acyl-CoA**

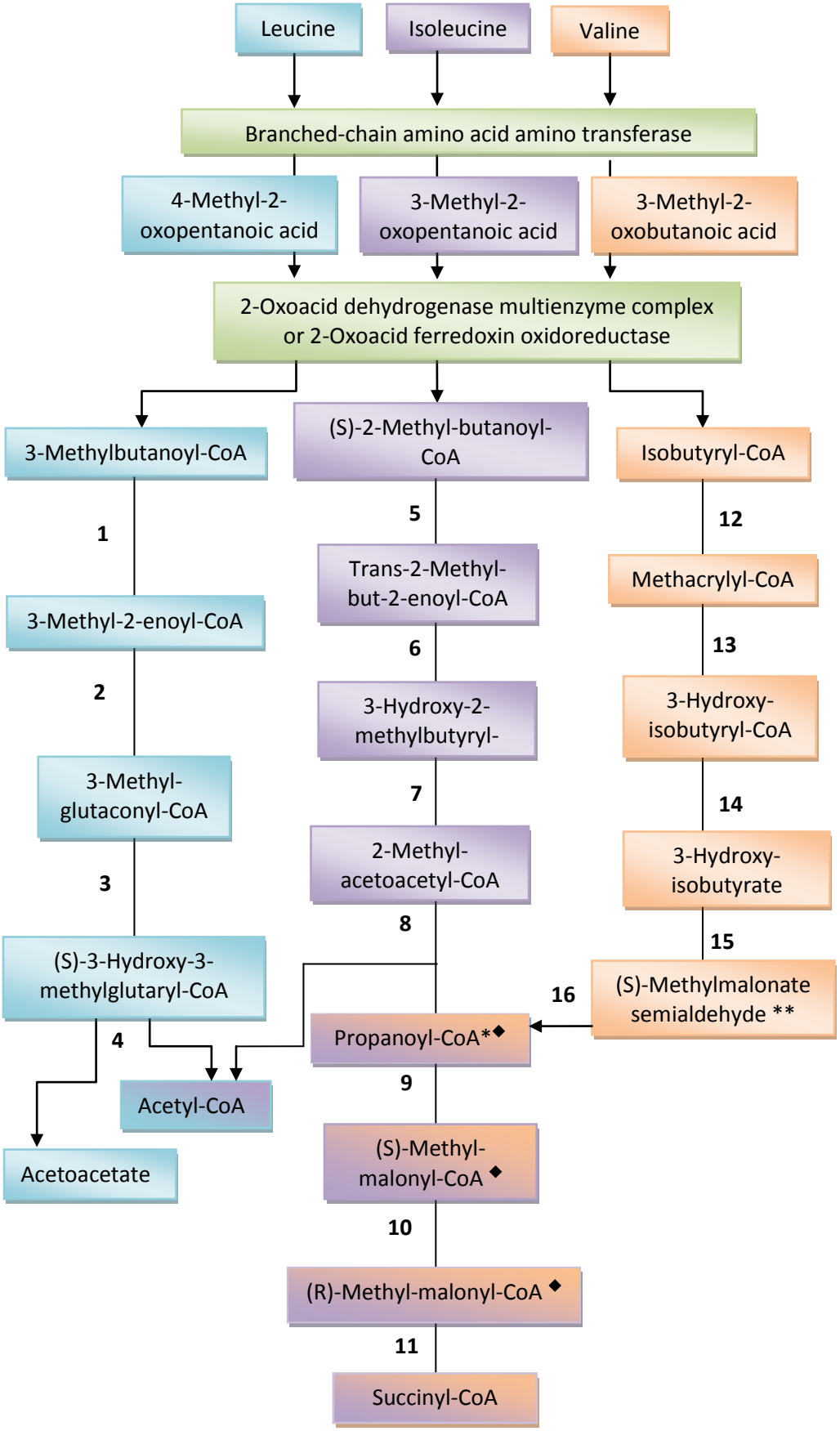


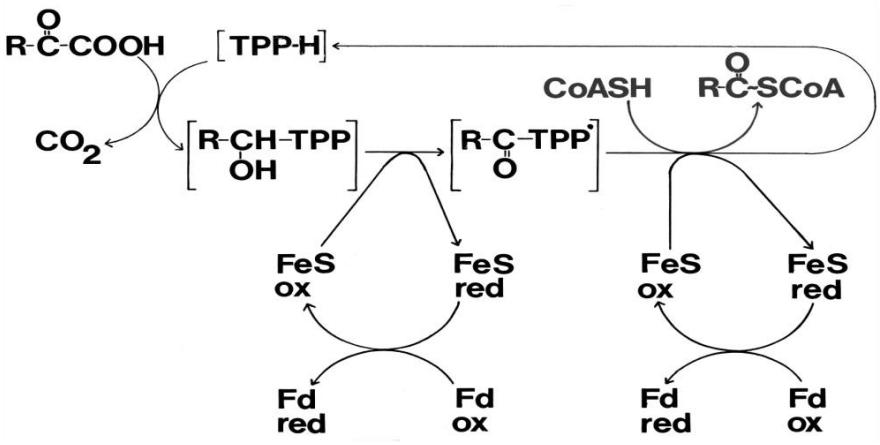

**Figure 1.5 Flow diagram of the metabolism of branched-chain amino acids to acetyl-CoA, succinyl-CoA, and acetoacetate.** (Details on the next page)

**Figure 1.5 Flow diagram of the metabolism of branched-chain amino acids to acetyl-CoA, succinyl-CoA, and acetoacetate**. The conversion of the various 2-oxoacids to their corresponding acyl-CoAs can be catalysed by a 2-oxoacid dehydrogenase multienzyme complex or a 2-oxoacid ferredoxin oxidoreductase. The enzyme responsible for catalysing each step, unless otherwise labelled, are numbered as following: [1] Isovaleryl-CoA dehydrogenase; [2] Methylcrotonyl-CoA carboxylase; [3] γ-Carboxygeranoyl-CoA hydratase; [4] hydroxymethylglutaryl-CoA lyase; [5] Acyl-CoA dehydrogenase; [6] 3-Hydroxyacyl-CoA dehydrogenase; [7] 3-hydroxyacyl-CoA dehydrogenase; [8] Acyl-CoA thiolase; [9] Propionyl-CoA carboxylase; [10] Methylmalonyl-CoA epimerase; [11] Methylmalonyl-CoA mutase; [12] Butyryl-CoA dehydrogenase; [13] 3-Hydroxyacyl-CoA dehydrogenase; [14] 3-Hydroxyisobutyryl-CoA hydrolase; [15] 3-Hydroxyacyl-CoA dehydrogenase; and [16] methylmalonate-semialdehyde dehydrogenase. * and ** indicate that the product can be diverted into propanoate and pyrimidine metabolism respectively; ◆ indicates that the products can be diverted into biosynthetic pathways.

### 1.2.3 Ferredoxin oxidoreductase

The conversion of 2-oxoacids to their corresponding acyl-CoAs in archaea and anaerobic bacteria is catalysed by a family of 2-oxoacid FOR enzymes including the pyruvate FOR (EC 1.2.7.1), 2-oxoglutarate FOR (EC 1.2.7.3) and the branched-chain 2-oxoacid FOR (EC 1.2.7.7). These enzymes can be promiscuous, exhibiting catalytic activity with a range of 2-oxoacids; for example, the 2-oxoglutarate FOR from *Sulfolobus solfataricus* P1 is capable of also utilising pyruvate and 2-oxobutyrate (Park et al., 2006).

The catalytic mechanism utilised by these enzymes is shown in Figure 1.6, as initially determined for the 2-oxoacid FOR from *Halobacterium halobium* by Kerscher and Oesterhelt (1981). The 2-oxoacid binds to the enzyme and is decarboxylated, leaving a reduced enzyme-bound iron-sulphur cluster, this is then reoxidised by ferredoxin and a TPP-acyl intermediate radical. Co-enzymeA (CoA) binds to the enzyme, resulting in the transfer of the acyl group from the TPP cofactor and the transfer of a second electron via an iron-sulphur centre onto a second ferredoxin molecule (Kerscher and Oesterhelt, 1981).




**Net reaction: Pyruvate + $Fd_{ox}$ + CoA →Acetyl-CoA + $CO_2$ + $Fd_{red}$ +$2H^+$**

**Figure 1.6 A reaction scheme of the 2-oxoacid ferredoxin oxidoreductase enzyme used by archaea and anaerobic bacteria**. The enzyme utilises a thiamine pyrophosphate (TPP) cofactor to decarboxylate the 2-oxo-acid, and the resulting acyl-group is subsequently transferred to Co-enzymeA (CoASH) with the concomitant transfer of the reducing equivalents via iron-sulphur centres (FeS) to ferredoxin (Fd). Diagram was provided by Professor M.J. Danson, University of Bath.

The FORs comprise multiple copies of between two and eight polypeptides; the number and combination of these polypeptides varies depending on the organism. Thermophilic organisms tend to contain FORs comprising four subunit types, for example the *Thermococcus profundus* branched-chain 2-oxoacid FOR is a hetero-octamer $(\alpha\beta\gamma\delta)_2$. On the other hand, mesophilic FORs tend to be dimers such as the *Desulfovibrio africanus* (Figure 1.7) FOR that has two large A subunits. The large mesophilic subunits have possibly evolved by the fusion of the genes encoding the various smaller polypeptides (Ozawa et al., 2005).

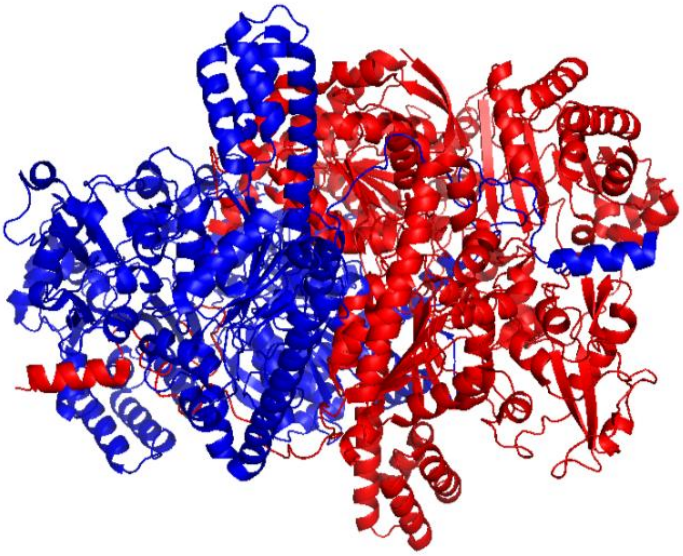


**Figure 1.7 Crystal structure of the *Desulfovibrio africanus* pyruvate ferredoxin oxidoreductase**. The ribbon dimer is composed of two identical polypeptides coloured red and blue, PDB access code 1B0P (Chabrière et al., 1999).

### 1.2.4 The 2-oxoacid dehydrogenase multienzyme complex

The 2-oxoacid dehydrogenase multienzyme complexes (OADHC) catalyse the conversion of 2-oxoacids to their corresponding acyl CoAs, along with the release of $CO_2$; unlike the FORs, the reducing equivalents are transferred onto $NAD^+$ rather than ferredoxin. It has been suggested that the OADHCs evolved after the evolution of oxidative phosphorylation, due to the presence of OADHCs only in aerobic organisms, whereas the FORs are utilised by anaerobic bacteria and archaea (Kerscher and Oesterhelt, 1982). It was originally thought that both aerobic and anaerobic archaea utilise only the FOR enzymes to catalyse the conversion of 2-oxoacids into their corresponding acyl-CoAs; but the discovery of the OADHC genes in several archaeal species opens up the possibility that archaea utilise both systems (discusses later in Chapter 1.3).

Structurally the FORs and OADHCs are very different; the FORs are smaller and simpler than OADHCs, which can be up to 10 MDa in size. The OADHCs comprise multiple copies of three enzymes: E1, E2, and E3. The E2 enzyme forms the structural core of the complex, 8 or 20 E2 trimers form a cubic or dodecahedral hollow cage like structure, to which the E1 and E3 enzymes associate non-covalently (Figure 1.8) (Perham, 2000).

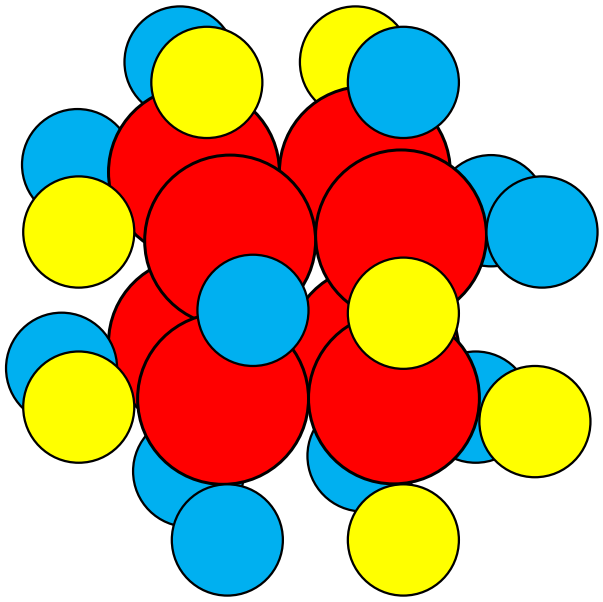
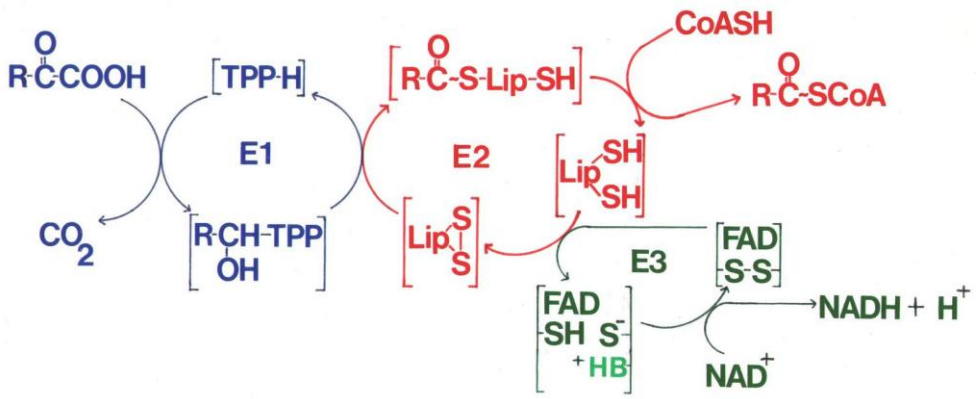


**Figure 1.8 The structure of a 24-mer 2-oxoacid dehydrogenase multienzyme complex**. The 8 E2 trimers (red circles) form the central core of the complex, and the E1 (yellow) and E2 (blue) enzymes then associate non-covalently with the E2 resulting in a 6 MDa complex.

The E1 enzyme is a 2-oxoacid decarboxylase. The 2-oxoacid is decarboxylated and then the acyl group is transferred onto the lipoic acid moiety attached to the end of the E2 swinging arm (E2 details in 1.2.4.2). The acyl group is then transferred from the lipoic acid onto CoA by the E2 dihydrolipoamide transacylase enzyme; this process leaves the lipoic acid moiety disulphide bond in its reduced form, this is then reoxidised and thus regenerated by the E3 dihydrolipoamide dehydrogenase enzyme.




**Net reaction: 2-oxoacid + CoA + $NAD^+$ → Acyl-CoA $CO_2$ + $NADH+H^+$**

**Figure 1.9 A generalised reaction scheme of the 2-oxoacid dehydrogenase multienzyme complex.** The three enzymes, E1 (blue), E2 (red) and E3 (green), catalyse consecutive reactions. Courtesy of Professor M.J. Danson (University of Bath)

### *1.2.4.1 E1 - 2-oxoacid decarboxylase*

The E1, 2-oxoacid decarboxylase, enzyme (EC 1.2.4.1, 1.2.4.2 or 1.2.4.4), catalyses the decarboxylation of a 2-oxoacid, along with the concomitant reductive acylation of an E2-lipoic acid moiety and the liberation of $CO_2$. The E1 enzymes are arranged as either homodimers ($\alpha_2$), or heterotetramers ($\alpha_2\beta_2$) (Figure 1.10). The homodimeric α-subunit (Figure 1.10A) is thought to have evolved from the fusion of the original α and β genes; thus, the *Pseudomonas putida* branched-chain OADHC E1α and β chains (Figure 1.10B) contain 410 and 339 residues respectively, in comparison to the *E. coli* PDHC homodimeric E1α that comprises 886 amino acids (Ævarsson et al., 1999, Arjunan et al., 2002). The *Tp. acidophilum* E1 and, based on gene structures, the E1 enzymes from all other archaea comprise E1α and E1β subunits; therefore, the structure and mechanism of only the heterotetramers will be discussed further.

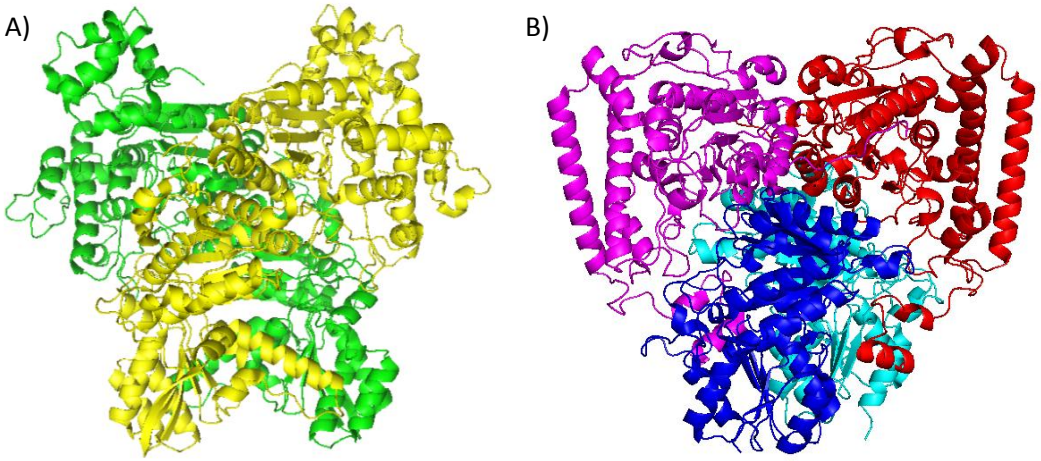



**Figure 1.10 The crystal structures of A) the homodimeric E1 from *E. coli* PHDC and B) the heterotetrameric E1 from *Pseudomonas putida* branched-chain OADHC.** The *E. coli* monomer ribbon diagrams are shown in yellow and green [PDB access code 1L8A (Arjunan et al., 2002)]. The *P. putida* $\alpha_2\beta_2$ ribbon diagram is shown with the α subunits coloured red and magenta and the β subunits coloured blue and cyan (PDB access code 1QS0 (Ævarsson et al., 1999)). Both images were created using PyMOL.

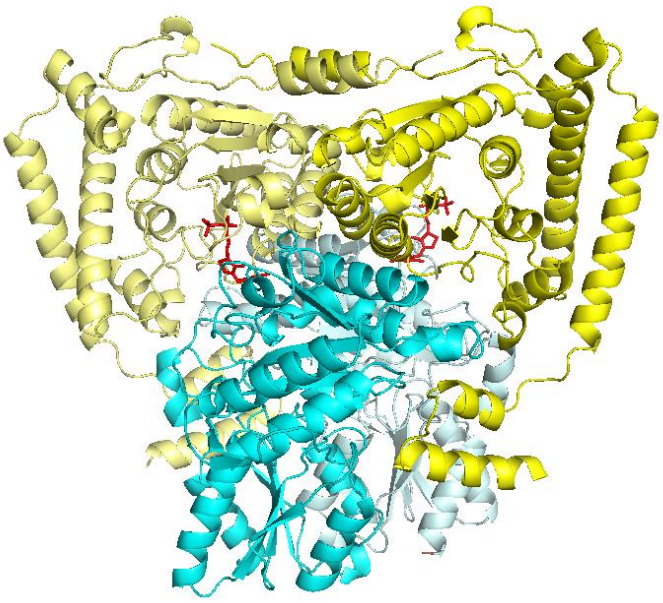
The structures of the E1α and β subunits mainly comprise α/β-fold domains. The α-subunit comprises one large α/β domain, a small C-terminal α-helical domain and an N-terminal extension; the β-subunit comprises two α/β domains connected by an extended loop region. The four polypeptides are arranged so that the two β-subunits are held in a vice like grip between the small C-terminal domains of the two α-subunits; the two E1α subunits are secured to each other by their N-terminal extensions (Figure 1.10) (Ævarsson et al., 1999). The E1 enzyme associates non-covalently with the E2 core of the OADHC, via the enthalpy driven, water-mediated interactions between the E1β subunits and the E2 peripheral subunit binding domain (PSBD). The binding of the PSBD to the E1 does not change the structure of the E1 and is not required for 2-oxoacid decarboxylase activity (Frank et al., 2005, Jung et al., 2002).

The E1, as the first enzyme in the OADHC, is responsible for conferring the 2-oxoacid substrate specificity of the complex; unlike the E2 and E3 enzymes which can accept exogenous lipoic acid, the E1 enzyme can only utilise lipoyl groups attached to its cognate E2 (discussed in Chapter 1.2.4.2/Figure 1.15). There are two active sites per E1$\alpha_2\beta_2$ located at the interface between the α- and β-subunits (Figure 1.11A); the active site containing the $Mg^{2+}$ and TPP cofactors is accessed via a hydrophobic tunnel (Ævarsson et al., 1999). The TPP cofactor is bound in the active site in a V-shaped formation; the phosphate end of the TPP molecule binds to the $Mg^{2+}$ and the E1α-subunit's TPP binding

motif (GDG(X)$_{20+}$-NN) (Hawkins et al., 1989); the aminopyrimidine end of the TPP binds to the E1β-subunit (Ævarsson et al., 1999). Forcing the TPP cofactor into a V-shaped formation, along with the abstraction of a proton by a nearby glutamate residue, acts to convert the cofactor into a highly reactive carbanion/ylide (Pei et al., 2008). The E1 catalyses the essentially irreversible decarboxylation and transfer of the acyl group by a two step ping-pong mechanism (Fries et al., 2003). The first reaction is the nucleophilic attack of the TPP carbanion on the 2-oxoacid, resulting in the liberation of $CO_2$ and the formation of an unstable acyl-TPP intermediate. The second reaction is the reductive acylation of the lipoic acid moiety, attached to the E2 lipoyl domain; it is this step that is rate limiting in the overall oxidative decarboxylation of the 2-oxoacids by an OADHC [Reviewed in Fries et al. (2003) and Pei et al. (2008)].

The X-ray crystal structure of the *Geobacillus stearothermophilus* PDHC E1$\alpha_2\beta_2$ shows asymmetry between the two αβ dimers, with one dimer in an "open" accessible conformation and the other in a "closed" confirmation; in the closed state, the entrance to the E1 active site is obscured by a surface loop (Frank et al., 2005). This asymmetry in the X-ray structure, along with the apparent increase in E1 activity after limited proteolysis, has led to the suggestion that the E1 active sites communicate with each other, allowing only one site to be functional at any one time (Chauhan et al., 2000). The two active sites are linked via a water-filled cavity and a chain of acidic residues (Figure 1.11B); it has been suggested that a "proton wire" allows the transfer of a proton from the activated TPP in one active site to the cofactor in the adjacent site, thus preventing activation of the second TPP (Frank et al., 2004). Mutation of *G. stearothermophilus* E1 Asp180 to Asn and Glu183 to Gln, two of the acidic residues located within the tunnel, reduced the rate of the E1 catalysed reaction (Frank et al., 2004), and similar studies on the *E. coli* PHDC E1$\alpha_2$ found that mutation of the acidic residues prevented transfer of information between the two active sites and reduced the binding of TPP in the correct orientation (Nemeria et al., 2010).

A)



B)

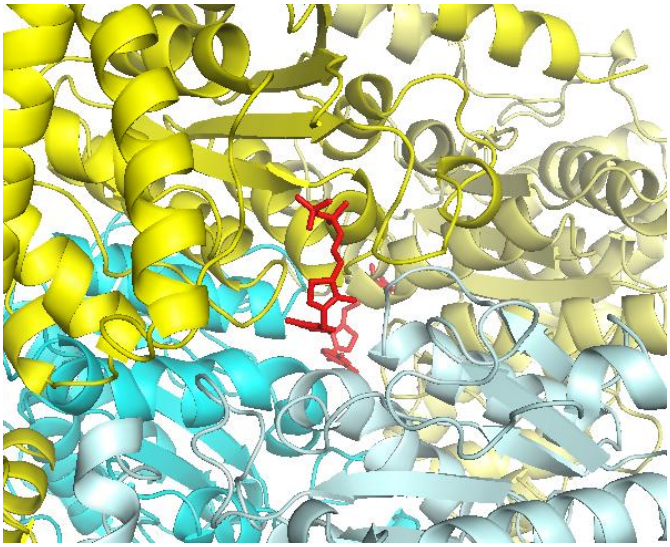


**Figure 1.11 The crystal structure of the *G. stearothermophilus* PDHC E1 heterotetramer.** A) E1$\alpha_2\beta_2$ ribbon diagram with the TPP cofactor shown in red; B) magnified view of the above structure rotated 90° to the right, showing the water-filled channel linking the two active site TPPs. The α- and β-subunits are shown in yellow and blue, respectively. PDB access code: 1W88.

### *1.2.4.2 E2 - Dihydrolipoamide transacylase*

The E2, dihydrolipoamide transacylase (EC 2.3.1.12 or 2.3.1.61), catalyses the transfer of the lipoamide-attached acyl group onto CoA, releasing acyl-CoA and reducing the lipoyl disulphide. Each E2 polypeptide comprises three separate domains connected by two flexible linker regions (Figure 1.12) (Perham, 2000):

- An N-terminal lipoyl domain (~80 residues) that transfers the intermediates between active sites via its covalently-attached lipoyl group,
- A peripheral subunit binding domain (PSBD) (~40 residues) that is responsible for the non-covalent association of the E1 and E3 enzymes with the E2 core
- A C-terminal catalytic domain (~230 residues) that has two roles: it contains the E2 acyltransferase active site, and this domain also forms the structural E2 core of the OADHC.

The flexible linkers play a crucial role, the N-terminal linker allows the lipoyl domain to move between active sites and thus transfer the reaction intermediates. The spatial arrangement of the E1 and E3 enzymes relative to the E2 core is controlled by the C-terminal linker region, this region adopts an extended conformation, resulting in a shell of E1 and E3 enzymes 75-90Å away from the E2 core (Lengyel et al., 2008) (Figure 1.12).

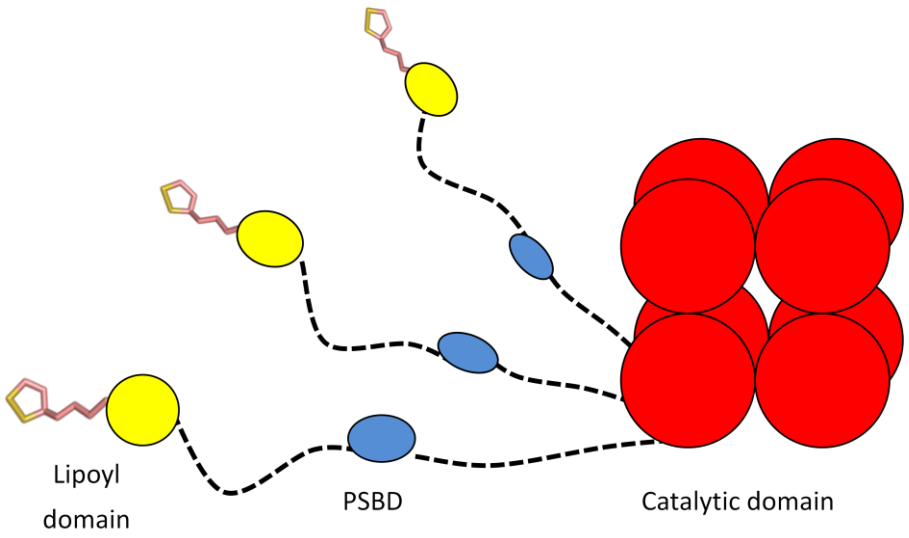



**Figure 1.12 Diagram of the domain structure of the E2 protein**. The eight catalytic domain trimers are represented in red, the PSDB (blue), lipoyl domain (yellow) and linker regions (black dashes) are shown for each of the monomers within one trimer.

The C-terminal E2 catalytic domain

There are several structures available for the E2 catalytic domain (E2cat) including *E. coli* succinyltransferase (PDB access code: 1E2O), *Geobacillus. stearothermophilus* and *Azotobacter vinelandii* acetyltransferase (1B5S and 1EAA, respectively), and *Pseudomonas putida* and *Bos taurus* branched-chain acyl-transferases (1QS0 and 2IHW). The E2cat monomers initially form trimers, and then either 8 or 20 trimers associate into a cubic 24-mer (Figure 1.13A) or a dodecahedral 60-mer (Figure 1.13B) respectively, depending on the organism and the complex. The E2 24- and 60-mers are both hollow cage-like structures comprising either square or pentagonal faces, with a large opening in the centre of each face, allowing free movement of substrates and products into and out of the interior of the E2 structure (Mattevi et al., 1992). The 24-and 60-mer structures are both Euclidean shapes, in which all subunits are in identical environments: the 24-mer displays octahedral 4,3,2 symmetry and the 60-mer displays icosahedral 5,3,2 symmetry; the only other form of cubic symmetry is tetrahedral 3,2 symmetry (Caspar and Klug, 1962). It is not possible to arrange more than 60 subunits into a perfectly symmetrical structure, in which all monomers are in identical environments. In the case of virus capsids which can comprise over 60 polypeptides, the structures are said to be quasi-symmetrical arrangements, the subunits inhabit slightly differing environments.

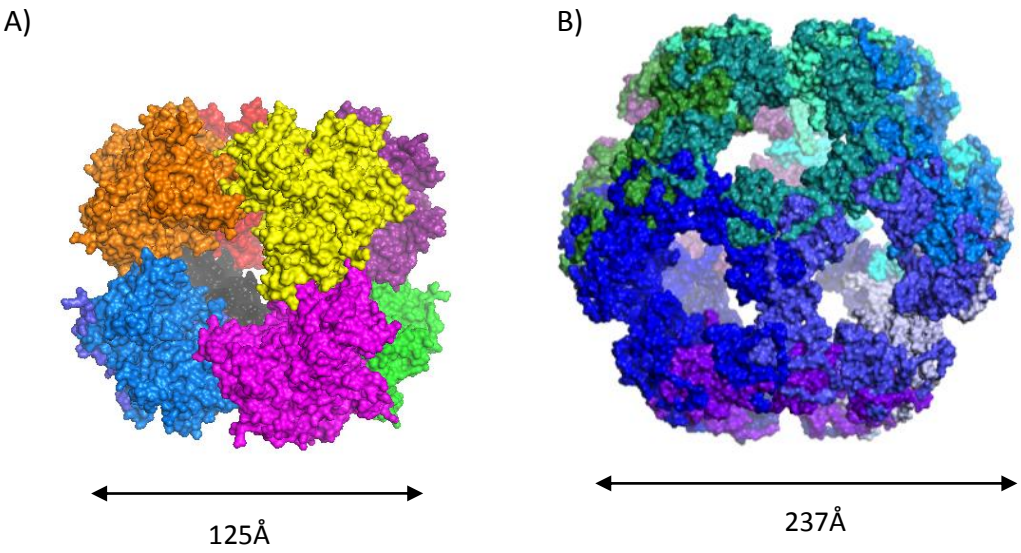



**Figure 1.13 A Surface representation of A) the cubic 24-mer, and B) the dodecahedral 60-mer.** The cubic 24-mer is the pyruvate dehydrogenase complex from *Azotobacter vinelandii* (PDB code: 1EAF) (Mattevi et al., 1992) and the 60-mer is the pyruvate dehydrogenase complex from *Geobacillus stearothermophilus* (PDB code: 1B5S) (Izard et al., 1999). The maximum diameter is shown below each structure (Image created with PyMOL).

The E2cat monomers have a structure similar to the chloramphenicol acetyltransferase enzyme (Knapp et al., 1998); they comprise a large β-sheet, surrounded by five α-helices and a C-terminal $3_{10}$ helix (Mattevi et al., 1992). The construction of trimers from the individual *A. vinelandii* E2 monomers is accompanied by the formation of 12 hydrogen bonds and the burial of 25% of the monomer surface area (Mattevi et al., 1992). The trimer structure of the 24- and 60-mers are identical, and it is the angle at which the trimers interact with one another that influences the overall structure. The inter-trimer interactions are mediated via a hydrophobic "anchor residue"; this residues, either a Met, Leu or Ile, is located a few residues in from the C-terminus of the protein on the $3_{10}$ helix (Izard et al., 1999); the hydrophobic side chain extends from one trimer into a hydrophobic pocket of a second trimer and *vice versa*. No correlation between the oligomeric nature of the E2 assemblies and the E2cat sequence have been identified; instead, it has been suggested that small changes in the interactions between the anchor residue and the side chains within the pocket may influence the angle of inter-trimer interactions, and thus the assembly of the E2 core (Izard et al., 1999). The importance of the anchor residue has been demonstrated by Peng *et al*. (2012), as the removal of the $3_{10}$ helix containing the Met 425 anchor residue from the *G. stearothermophilus* E2cat resulted in a protein that no longer assembled into a 60-mer.

The E2 transacylase active site is located within a ~30Å channel, situated at the intra-trimer monomer interfaces, and each of the three active sites contains residues contributed by two monomers. The channel traverses the trimers with an opening on both the inside and the outside of the assembled E2; the acylated lipoic acid enters the active site from the outside, and the CoA enters from the interior of the E2 cage (Mattevi et al., 1993). Binding of substrates to the *A. vinelandii* and *E. coli* E2 has no effect on the structure. However, there are differences between the bovine branched-chain E2 crystal structures with and without CoA bound, and it has been suggested that the eukaryotic E2 may have a substrate-gating mechanism (Kato et al., 2006). In the absence of CoA, a cluster of residues (Leu 93, Val 295 and Leu 353) prevent the acylated-lipoamide from entering the active site channel; however, upon binding of CoA to the active site a conformational change rotates these gating residues away from the channel entrance, thus exposing the lipoamide binding site. The role of this gating mechanism is possibly to increase the efficiency of the E2 enzyme, by only allowing the acylated-lipoamide to enter active sites that already have a CoA present. This adaptation may increase efficiency by reducing the chance of acyl groups being lost by hydrolysis in the absence of CoA and/or

to reduce the time that each lipoamide spends bound to an E2 waiting to be deacylated (Kato et al., 2006).

The E2 is thought to catalyse the transfer of the acyl group from the lipoic acid moiety onto CoA via a mechanism similar to that used by the chloramphenicol acetyltransferase (Guest, 1987). The CoA thiol group is first deprotonated by a histidine imidazole side chain. The negatively-charged $CoAS^-$ then attacks the acylated lipoic acid, resulting in a tetrahedral intermediate that is stabilised by a Thr or Ser residue, and correctly orientated by an Asp (Knapp et al., 1998); this intermediate then collapses, resulting in the formation of acyl-CoA and a reduced E2 bound lipoic acid.

Following the determination of the *G. stearothermophilus* E2 structure by Izzard et al. (1999), one branch of research has focused on the potential medical applications this protein may have. One possibility is to replace the PSBD and lipoyl domains with a viral protein or protein motifs, in order to create a vaccine for HIV (Domingo et al., 2001). The advantage of using the E2 to present antigens is that multiple viral protein motifs and immune cell activating proteins can be displayed on a single 60-mer, resulting in a synergistic immune response (Caivano et al., 2010). Another potential use of the E2 protein being investigated is the ability of the E2 catalytic domain to encapsulate a drug molecule whilst display targeting molecules on its outer surface (Dalmau et al., 2008). This research aims to add a solvent-accessible Cys residue to the interior surface of the E2cat protein, so that a drug molecule can be attached to the inside of the cage, thus encapsulating and sequestering it. The addition of a targeting protein to the surface could then target the encapsulated drug to a certain cell type. The next stage of this research was to modify the E2 structure by the addition of histidine residues at the trimer-trimer interfaces. The aim is to create an E2 capsule that sequesters the drugs until the protein is endocytosed; during endocytosis the acidification of the vesicle causes the imidazole side chains at the trimer-trimer interface to become positively charged, repel each other and trigger disassembly of the E2 protein along with the release of the sequestered drug. Currently, the pH-dependent disassembly of this E2 protein is unfortunately coupled with its aggregation, which may cause unwanted side effects (Dalmau et al., 2009).

The peripheral subunit binding domain (PSBD)

The PSBD is located between the N-terminal lipoyl domain and the C-terminal catalytic domain; this small domain, comprising 40-50 residues, is responsible for the non-covalent association of the E1 and E3 enzymes with the OADHC E2 core. This domain and its role in OADHC assembly, were determined by limited proteolysis of the E2 in the presence of the E1 and E3 enzymes; gel filtration of the resulting protein fragments found that a small E2 domain remained attached to the E1 and E3 proteins (Duckworth et al., 1982). The structure of this domain was solved by NMR, using a 43-residue synthetic peptide based on the *G. stearothermophilus* sequence. The PSBD structure is a compact domain (25Å x 20Å x 20Å) comprising two short α-helices and one $3_{10}$ helix (Kalia et al., 1993). The structure of the recombinantly expressed *G. stearothermophilus* PSBD in complex with the E3 has since been determined by X-ray crystallography (Figure 1.14) (Allen et al., 2005). The PSBD is one of the smallest known independently-folding protein domains with a stable tertiary structure; this along the short sequence allowing for synthetic production and thus the ability to incorporate non-natural amino acids, has meant that the PSBD has been an ideal model for studying the energetics and physical processes involved in protein folding and unfolding (Sharpe et al., 2008, Spector et al., 1998, Ferguson et al., 2005).

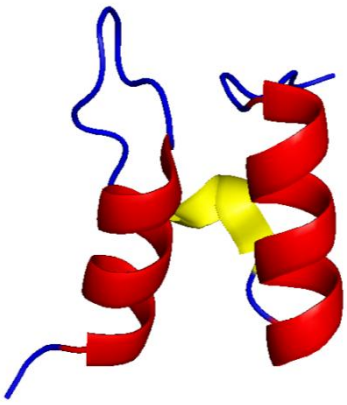


**Figure 1.14 The X-ray crystal structure of the *G. stearothermophilus* PSBD domain.** The two α-helices and the $3_{10}$ helix are shown in red and yellow respectively; the loop regions are shown in blue (Allen et al., 2005). (PDB access code: 1W3D).

The E1 and E3 enzymes bind competitively to the PSBD. The $E1\alpha_2\beta_2$ interaction is mediated via the β-subunits; the *G. stearothermophilus* E1β has been shown to bind to the PSBD without the E1α subunit. The E1 tetramers and E3 dimers have both been shown to bind separately to the PSBD in a stoichiometry of 1:1 (Lessard and Perham, 1995, Hipps et al., 1994), with the E1 binding more strongly than the E3 (Jung et al., 2003).

The biophysical data and the X-ray crystallographic structures of the PSBD, bound to either the E1 or E3 (Figure 1.15), confirm that E1 binds to only one PSBD at the E1β interface (Frank et al., 2004), and the E3 dimers also bind to only one PSBD at the dimerisation interface (Mande et al., 1996). Mutagenesis studies of the *G. stearothermophilus* PSBD have identified two Arg residues (Arg 135 and 156) as vital in the formation of interactions between the PSBD and the E1 or E3. Met 131 has also been identified as crucial for the formation of the E1-PSBD interactions but is not required for E3 binding (Jung et al., 2003); these data suggest that the PSBD has two overlapping binding sites for the E1 and E3 enzymes.

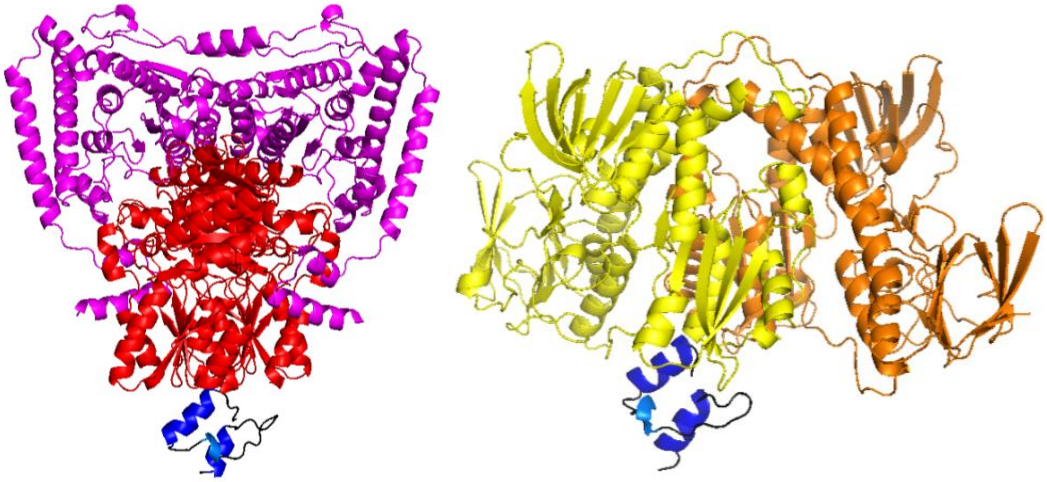


**Figure 1.15 The crystal structures of the *G. stearothermophilus* PDHC PSBD bound to A) the E1$\alpha_2\beta_2$ and B) the E3$\alpha_2$ enzymes shown as ribbon diagrams.** The PSBD α helices and $3_{10}$ helix are shown in dark and light blue, respectively. A) The E1 α- and β-subunits are shown in magenta and red, respectively. B) The two monomers of the E3$\alpha_2$ structure are shown in yellow and orange. E1 PDB access code: 1W88 (Frank et al., 2004); E3 PDB access code: 1EBD (Mande et al., 1996).

The situation in eukaryotes is slightly different, as only the E1 binds to the E2-PSBD; the E3 binds instead to an E3-binding protein (E3BP). The E3BP comprises a single lipoyl domain, an E3-binding domain and a non-catalytic core. The non-catalytic E3BP core domain, along with the E2 catalytic domain, assemble into a 60-mer; the exact stoichiometry *in vivo* is not clear but a ratio of 48:12 E2:E3BP has been suggested (Brautigam et al., 2009). It has been proposed that the human E3BP integrates into the OADHC by replacing some E2 subunits (Yu et al., 2008). The mammalian E2 and E3BP are closely related homologues; however, the E2 active site His is replaced in the E3BP by a Ser, the E3BP

also has only one lipoyl domain unlike the E2 which has three (Hiromasa et al., 2004). This alternative eukaryotic arrangement, may allow for increased control of the stoichiometry of E1 and E3 enzymes in the assembled OADHC, in a more subtle manner than by simply controlling expression levels of the two proteins.

The N-terminal lipoyl domain

The N-terminal lipoyl domain has a lipoic acid moiety covalently attached to a lysine residue; this chemical modification allows the OADHC reaction intermediates to bind to this domain and be transferred between the active sites. The number of lipoyl domains can vary between one and three depending on the complex and the organism: *Tp. acidophilum* OADHC has one, *Neisseria meningitidis* and *Alcaligenes eutrophus* PDHCs have two and the *A. vinelandii* and *E. coli* PDHCs have three (Perham, 2000).

The structure of this domain has been determined by NMR. The lipoyl domain comprises two β-sheets; a lipoic acid moiety is covalently attached to a lysine residue that is located on a exposed surface loop flanked by moderately-conserved Asp and Ala residues (Figure 1.16) (Reche and Perham, 1999, Perham, 2000). It has been shown that the *G. stearothermophilus* E1 enzyme recognises the lipoylated lipoyl domain preferentially over either a non-lipoylated domain or free lipoic acid; thus, the E1 active site interacts with the β-turn and surface loop of the lipoyl domain, in addition to the lipoic acid. The E3 enzyme also preferentially accepts lipoic acid tethered to a lipoyl domain, but in this case the active site interacts only with the lysine and two adjacent amino acids (Figure 1.15) (Fries et al., 2007).

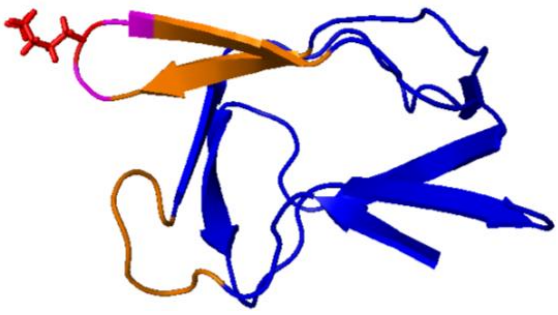


**Figure 1.16 The *G. stearothermophilus* PDHC E2 lipoyl domain ribbon diagram.** The lysine residue to which the lipoic acid is attached is highlighted in red. The magenta regions are the residues either side of the Lys that the E1 and E3 active sites recognise; the orange regions are the residues recognised by the E1 active site in addition to the pink residues PDB access code: 1LAB (Fries et al., 2007). Image created in PyMOL.

Lipoylation of the E2 lipoyl domain lysine can occur via two pathways in *E. coli* depending on the availability of exogenous lipoic acid. When lipoic acid is exogenously available, LplA (EC: 2.7.7.63), an ATP-dependant lipoate protein ligase, catalyses the formation of an amide bond between the lipoic acid and the Lys ε-amino group (Morris et al., 1994). A second pathway is utilised when lipoic acid is not freely available; in this case, octanoic acid, produced by the fatty acid biosynthesis pathway, is attached to the lipoyl domain by LipB, a lipoyl (octanoyl) transferase (EC 2.3.1.181). The octanoic acid is then converted via octanoyl-ACP into lipoyl-ACP by LipA, a lipoyl synthase (EC 2.8.1.8) (Morris et al., 1994).

The flexible linker region connecting the lipoyl domain and the PSBD lets the lipoyl domain to move freely between the active site of the OADHC (Perham, 1991); allowing the covalently attached reactive intermediates to be stabilised and efficiently transported between active sites. The efficiency of the OADHC is also increased by the lipoyl domains ability to transfer intermediates onto other lipoyl domains, and thus channel substrates around the complex; at low substrate concentrations this ability allows the sequential OADHC reactions to occur, even if the substrates were initially dispersed around the complex (Danson et al., 1978).

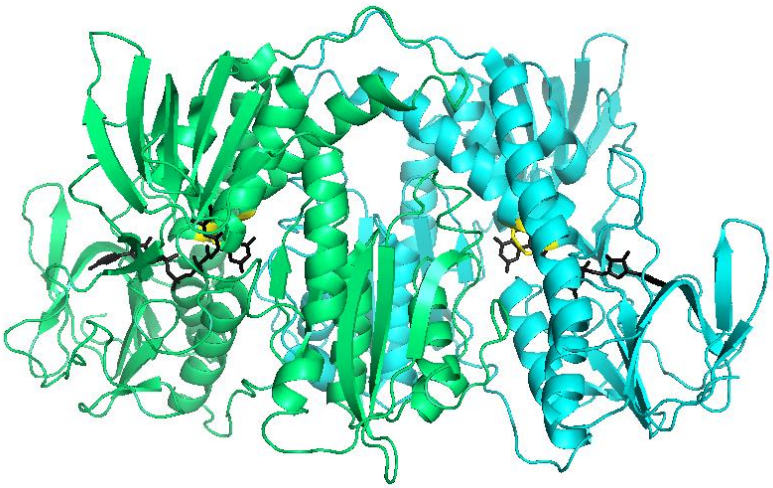
#### *1.2.4.3 E3 - Dihydrolipoamide dehydrogenase*

The E3 dihydrolipoamide dehydrogenase catalyses the following reaction:

$$E2\text{-}Lip(SH)_2 + NAD^+ \leftrightarrows E2\text{-}LipS_2 + NADH + H^+$$

The homodimeric E3 enzyme is part of a family of dehydrogenases including the glutathione, trypanothione and mercuric reductases, all of which have a similar overall structure and have a reaction mechanism that utilises an active-site disulphide bond. The *A. vinelandii* E3 structure has been determined to a resolution of 2.2Å by Mattevi et al. (1991a); each E3 monomer comprises 4 domains: the $FAD^+$ and $NAD^+$ binding domains (amino acids 1-150, and 151-280 respectively), a central domain (281-349) and an interface domain (350-466). The two active sites are located at the interface between the two dimers, with each monomer contributing residues to both sites. The active-site disulphide bond is formed between Cys48 and Cys53 (Figure 1.17). The two Cys residues are located on an α-helix with an extended turn; the distortion in the α-helix brings the two sulphurs to within 4.2Å, rather than 8.8Å without the distortion, allowing the formation of the crucial active-site disulphide bond (Mattevi et al., 1991a).

A)



B)

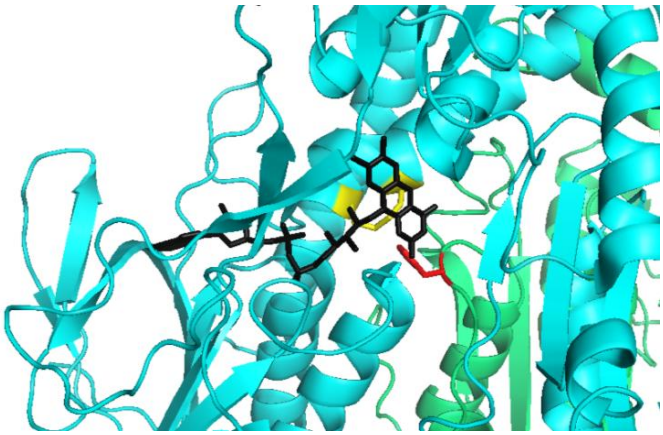


**Figure 1.17 The structure of the homodimeric *A. vinelandii* E3**. The two subunits are shown in blue and green in Image A. Image B is a magnified view of the active site of the blue monomer; the FAD cofactor is shown in black, the active-site disulphide between Cys48 and Cys53 is shown in yellow, and the catalytic His450', contributed by the green subunit, is shown in red. PDB access code 3LAD (Mattevi et al., 1991a) (Image created with PyMOL).

The lipoamide dehydrogenation reaction occurs in two steps: the oxidation of the lipoic acid moiety by the E3 enzyme, followed by the oxidation of the enzyme by $NAD^+$ (Thorpe and Williams Jr, 1981):

Reaction 1: $E + Lip(SH)_2 \leftrightarrows EH_2 + Lip(S)_2$

Reaction 2: $EH_2 + NAD^+ \leftrightarrows E + NADH+H^+$

The reduced lipoamide enters the active site via a narrow channel, where it reduces the enzyme's disulphide bond (Cys48-Cys53). Cys48 and the active site His450' are protonated, while the negative charge on Cys53 is stabilised by the formation of a charge transfer complex with the enzyme bound $FAD^+$. The oxidised lipoamide then leaves the E3 active site and, in the second part of the reaction, the reducing equivalents are transferred onto $NAD^+$, thus regenerating the enzyme and releasing NADH + $H^+$ (reviewed in Mattevi et al. (1991a)).

## 1.3 The OADHCs in Archaeal carbon metabolism

### 1.3.1 Evidence of OADHCs in Archaea

Archaea were originally thought to utilise only FORs to convert all the 2-oxoacids to their corresponding acyl-CoAs (Kerscher et al., 1982); however evidence has emerged that aerobic archaea may utilise both FORs and OADHCs (Smith et al., 1987).

#### *1.3.1.1 The detection of dihydrolipoamide dehydrogenase activity in archaeal cell lysates*

The first indication that aerobic archaea may utilise an OADHC, in addition to the FOR, was the detection of dihydrolipoamide dehydrogenase activity in cell lysates from *Halobacterium halobium*, a halophilic archaeon, and *Tp. acidophilum*, a thermoacidophilic archaeon (Danson et al., 1984, Smith et al., 1987). The activity detected in *H. halobium* was confirmed to be as a result of a dihydrolipoamide dehydrogenase, rather than a promiscuous glutathione reductase as the two activities could be separated by gel filtration (Smith et al., 1987). The presence of dihydrolipoamide dehydrogenase activity and the detection of lipoic acid, an essential cofactor for OADHC activity, by GC-mass spectrometry, in *H. halobium* (Pratt et al., 1989) led to the hypothesis that some species of archaea may possess an OADHC in addition to the FORs. E3 activity has since been detected in several species of aerobic archaea, but this has not been coupled with the detection of OADHC or E1 activity in any archaeal cell lysate.

#### *1.3.1.2 Identification OADHC genes archaea*

The E3 gene was identified in the halophile *Haloferax volcanii*; upstream sequencing of this gene identified the other three OADHC genes, E1α, E1β and E2, arranged in an operon-like structure. *H. volcanii* messenger RNA was screened for a region containing the E3 sequence, and a single 5.2kb mRNA strand was identified (Al-Mailem et al., 2008); this size closely matches the 5.4kb expected if the putative OADHC genes are arranged as an operon and are translated as a single mRNA molecule (Jolley et al., 2000).

The recent sequencing and annotation of the genomes of several species of archaea have identified the OADHC genes in aerobic organisms from mesophilic, halophilic and thermophilic environments (Figure 1.18).

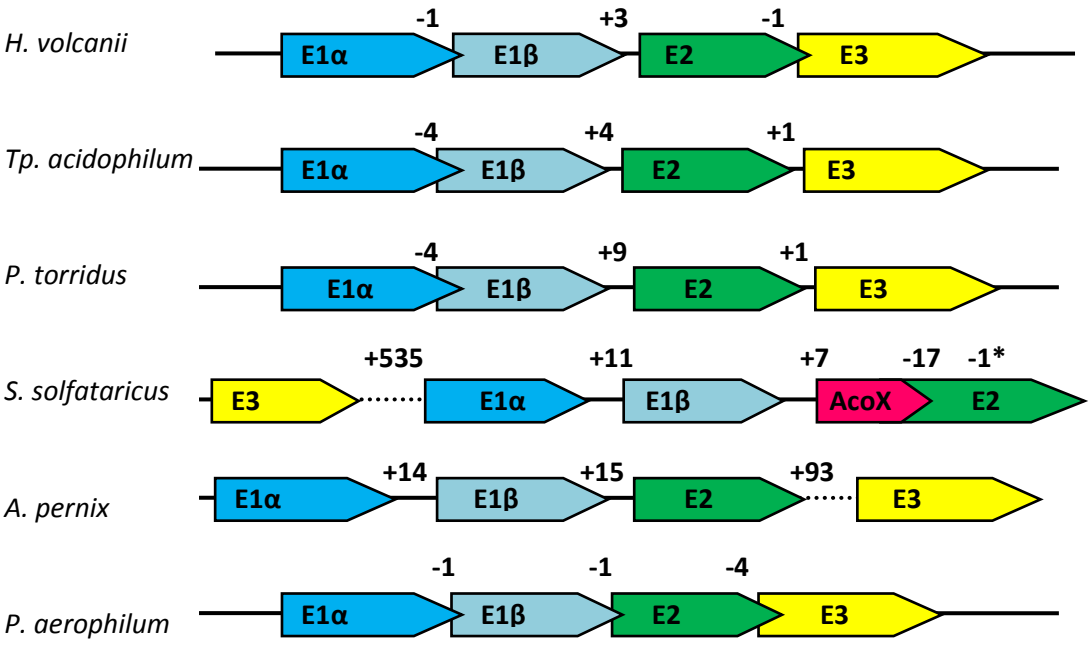



**Figure 1.18 Putative OADHC gene arrangements in aerobic archaea.** The genes identified as encoding the components of a putative OADHC are shown by arrows indicating the direction of the open reading frame, and the distances in base pairs between the genes are shown above. The -1* in the middle of the *S. solfataricus* E2 gene indicates that this gene contains a -1 base-pair frame-shift. The AcoX gene located within the *S. solfataricus* operon encodes a protein that may be involved in acetoin metabolism.

Transcriptomic studies of *Pyrobaculum aerophilum* have noted that the OADHC genes are transcribed, and that transcription appears to be regulated by the respiratory growth conditions of the organism (Cozen et al., 2009). A proteomics study of large complexes (>1 MDa) in *Tp. acidophilum* identified the E1 and E2 proteins but not the E3, (Sun et al., 2007), implying that the OADHC genes are transcribed and translated but under these conditions the E3 does not appear to associate with the E2 core.

In conclusion, there is a mounting body of evidence implying that aerobic archaea utilise the OADHCs; this evidence includes the presence of genes encoding the OADHC proteins, E1α, E1β, E2, and E3, along with the proteins required for E2 lipoylation, and that these genes appear to be transcribed and translated, with E3 activity being detected in several cases. The only result contradicting these findings, is that as yet no OADHC activity has been detected in cell lysates of the archaea.

### 1.3.2 Characterisation of the OADHC from *Tp. acidophilum*

The four complex genes, E1α, E1β, E2 and E3, were identified in the *Tp. acidophilum* genome along with a possible LplA gene (Section 1.3.3), suggesting that this organism has the potential to produce a functional OADHC. The OADHC genes were cloned and expressed by Heath (2006) and Aass (2003). The E1α and E1β assembled into an $\alpha_2\beta_2$ tetramer which had activity with branched-chain 2-oxoacids and to a lesser extent, pyruvate (Heath et al., 2004). The recombinantly expressed E2 was soluble and approximately 50% of the polypeptides were lipoylated by the *E. coli* cells when the medium was supplemented with lipoic acid; the E2 assembled into a large structure predicted to be a 24-mer by sedimentation velocity ultra-centrifugation. The E3 gene was recombinantly expressed in *E. coli*, resulting in a dimeric protein with an equimolar concentration of $FAD^+$ to E3 polypeptides (Aass, 2003). The purified recombinant *Tp. acidophilum* OADHC enzymes were found to assemble into a large 5 MDa multienzyme complex, which was capable of catalysing the oxidatative decarboxylation of branched-chain 2-oxoacids and pyruvate (Heath et al., 2007). These data prove that even though OADHC activity has not been detected in *Tp. acidophilum* cell extracts, the OADHC genes encode functional proteins that can assemble into a branched-chain OADHC.

### 1.3.3 Identification of an E2 lipoylation pathway in *Tp. acidophilum*

A search of the *Tp. acidophilum* genome for LplA or LipA/LipB genes, encoding the enzymes required to lipoylate the E2, identified a putative LlpA gene (Ta0514) (McManus et al., 2006); the LlpA protein was recombinantly expressed but lipoate protein ligase activity was not detected (PDB access code: 2C7I). The *Tp. acidophilum* and *E. coli* LplA protein sequences align well with each other, except *Tp. acidophilum* LplA is 90 amino acids shorter than the *E. coli* LplA. When the structure of the two LplAs were superimposed, the missing region equated to the *E. coli* C-terminal domain, a small domain that is separated from the rest of the protein by a flexible linker (McManus et al., 2006). A BLASTP search of the *Tp. acidophilum* genome for the missing 90 residues identified a short gene upstream of the putative LplA sequence (Ta0513m). Thus, *Tp. acidophilum* appeared to possess genes for separate N- and C-terminal LplA domains that might constitute a novel bipartite LplA enzyme, (Posner, 2009). The two proteins have been recombinantly expressed in *E. coli* and shown to associate with each other to give a catalytically-active LplA enzyme (Posner et al., 2009). The *Tp. acidophilum* genome

therefore contains the proteins required to lipoylate the E2 enzyme in the presence of exogenous lipoic acid. The presence of the bipartite LplA , E1, E2, and E3 genes suggests that even though OADHC activity has never been detected in cell lysate, it must have a role, as archaea do not maintain unnecessary genes (Csűrös and Miklós, 2009).

## 1.4 Aims of the project

To date studies of enzyme thermostability and thermoactivity have mainly been performed on individual isolated enzymes; however, this is not truly representative of the conditions within a cell, where a large number enzymes and proteins may be closely associated with one another. This project aimed to extend these studies from small oligomeric enzymes to a multienzyme complex, namely the OADHC from the thermophilic and hyper-thermophilic archaea by characterising the thermostability and thermoactivity of the *Tp. acidophilum* individual OADHC components and the fully assembled OADHC.

The E2 enzyme forms the structural core, around which the E1 and E3 enzymes associate; the thermostability of the E2 is fundamental to the stability of the whole multienzyme complex. Due to the flexible multi-domain nature of the E2 protein's structure, the whole E2 is unlikely to crystallise; therefore, the second aim of this project was to clone and express the catalytic domain of the E2, in order to determine a high-resolution X-ray crystallographic structure.

The final aim was to study other thermophilic and hyper-thermophilic OADHCs, in order to answer questions such as: what are the substrate specificities of the different complexes in the different species of archaea, how does this affect the possible physiological roles of the OADHCs and finally what is the evolutionary history of these complexes in archaea?

# Chapter 2: Material and Methods

In all cases, unless otherwise stated, laboratory reagents were supplied by Sigma-Aldrich Company Ltd. (Poole, U.K.), Melford Laboratories Ltd. (Ipswich, U.K.), or Fisher Scientific (Loughborough, U.K.), and were of the highest grade commercially available.

## 2.1 Microbiology techniques

### 2.1.1 *Escherichia coli* strains

Cloning experiments were carried out using *E. coli* JM109 cells, an endonuclease (*endA*) and recombinase (*recA*) deficient strain. The following modified *E. coli* strains were used for protein expression: *E. coli* BL21, *E. coli* Rosetta (DE3) pLysS (both Novagen, Merck, U.K.) and Arctic Express™ (Stratagene, Agilent Technologies, U.K.); these strains utilise the IPTG-inducible lac promoter to drive the expression of the T7 RNA polymerase, and the consequent over-expression of the target protein. The *E. coli* Rosetta strain contains an additional pRARE plasmid that supplies extra tRNAs, not commonly used by *E. coli*: AUA, AGG, AGA, CUA, CCC, and GGA. The ArcticExpress™ strain was utilised when the target protein was consistently expressed in an insoluble form by the other two strains. This strain contains the psychrophilic Cpn10 and Cpn60 chaperonins from *Oleispira antarctica*; these chaperonins are able to function between 4-12°C, assisting protein folding and thus increasing the overall yield of correctly-folded soluble protein.

### 2.1.2 Growth of *Escherichia coli*

Unless otherwise stated, all cultures were grown in 500ml Luria Bertani (LB) medium (1% (w/v) tryptone, 1% (w/v) NaCl, 0.5% (w/v) yeast extract and pH 7.0). The LB medium was supplemented with antibiotics as recommended by the pET vector system, Promega and Stratagene manuals (carbenicillin 50µg/ml, kanamycin 30µg/ml, chloramphenicol 34µg/ml, or 20µg/ml gentamycin).

### 2.1.3 Production of electrocompetent *E. coli* cells

To produce the electrocompetent *E. coli* JM109 and Rosetta cells, 5ml YENB medium (0.75% (w/v) yeast extract, 0.8% (w/v) nutrient broth, adjusted to pH7.5), supplemented with the required antibiotics, was inoculated with a single colony and incubated for 15 h

at 37°C; 2ml of this starter culture were then inoculated into 500ml YENB medium and incubated at 37°C, shaking at 200rpm, until the $A_{600}$ reached 0.8-1.0. The cells were then harvested by centrifugation at 13400 x g, at 4°C for 10 min. The cell pellet was then gently resuspended in 25ml cold $dH_2O$ and centrifuged again; this wash step was repeated with another 25ml $dH_2O$, and 30ml 10% (v/v) glycerol. The final cell pellet was resuspended in 900µl 10% (v/v) glycerol, and then divided into 40µl aliquots, snap-frozen and stored at -80°C.

## 2.2 Molecular biology techniques

The following molecular biological instructions were followed to clone and express a target gene in *E. coli*:

1. PCR amplify the required gene
2. Determine PCR products size and purity by agarose gel electrophoresis
3. Purify PCR product from agarose gel
4. A-tail the PCR product and ligate into pGem-T Easy vector
5. Transform the pGem-T Easy vector into *E. coli* JM109 for blue/white screening
6. Select several white *E. coli* colonies, and extract the plasmid
7. Confirm the presence and size of the insert in pGem-T Easy by restriction digestion using specific enzyme combinations
8. Sequence the selected pGem-T Easy plasmids to confirm gene sequence
9. Ligate the digested gene from pGem-T Easy into the required pET vector, that has been previously digested using the same restriction enzyme combination
10. Transform the pET vector into *E. coli* JM109
11. Select *E. coli* colonies to screen for the target gene using a colony PCR
12. Extract the pET vector plasmid from positive colonies
13. Confirm the positive colony PCR result by restriction digestion of the pET vector
14. Sequence the pET vector to ensure the correct gene is present without any mutations
15. Transform the pET vector into an *E. coli* expression strain for protein expression

### 2.2.1 Polymerase chain reaction

#### *2.2.1.1 DNA amplification*

Unless otherwise stated, all PCR reactions were performed using Phusion® polymerase PCR kits (New England Biolabs) following the conditions recommended by the supplier; a standard PCR mixture contained: 1x High Fidelity buffer, 0.2-1.0µM primers, 6ng/50µl template DNA, 0.02U/µl Phusion DNA polymerase, and 200µM dNTPs (dATP, dCTP, dGTP, dTTP) (New England Biolabs, Ipswich). The reactions were carried out in thin-walled Eppendorf PCR tubes in an Eppendorf Maser cycler (Eppendorf, Germany). A standard cycle is shown in Table 2.1. All primers were produced by Invitrogen (U.K.).

| | | Temperature (°C) | Time |
|---|---|---|---|
| x40 cycles | **Initial denaturation** | 98 | 30 sec |
| | **Denaturation** | 98 | 10 sec |
| | **Annealing** | Y | 30 sec |
| | **Elongation** | 72 | X min |
| | **Final elongation** | 72 | 10 min |

**Table 2.1 Temperatures and times of each stage of a standard PCR**. The elongation time (X) is dependent on the length of each DNA sample (30 sec per 1Kb). Temperature (Y) depends on annealing temperature of each pair of primers.

#### *2.2.1.2 Mutagenesis cycle*

Site-directed mutagenesis was utilised to mutate or delete short sequences of DNA in a plasmid. The method described in the Stratagene Quickchange® Site-Directed mutagenesis kit manual was used to design the primers and Phusion polymerase-PCR conditions (Table 2.2).

| | | Temperature (°C) | Time |
|---|---|---|---|
| X18 cycles | **Initial denaturation** | 98 | 30 sec |
| | **Denaturation** | 98 | 30 sec |
| | **Annealing** | 55 | 30 sec |
| | **Elongation** | 72 | X min |
| | **Final elongation** | 72 | 10 min |

**Table 2.2 The temperatures and times of each stage of a site-directed mutagenesis PCR**. The elongation time (X) is dependent on total size of the plasmid and gene (30 sec per 1Kb).

The methylated template DNA was removed, after the PCR, by a 60-minute incubation with *Dpn* I, leaving only the amplified plasmid containing the required mutation. The PCR mixture, containing the *Dpn* I enzyme, was then incubated at 80°C for 20 min to inactivate the *Dpn* I. The mutated plasmids were then purified, using the Wizard® SV Gel and PCR DNA Purification System (Promega) following the supplier's instructions, and transformed into *E. coli* JM109 for propagation. The plasmids were extracted from several *E. coli* colonies and sequenced by SourceBioscience plc (Oxford, U.K.) to ensure the correct mutation was present. The mutated plasmid was then transformed into the required *E. coli* protein expression strain (Chapter 2.2.5).

### 2.2.2 Agarose gel electrophoresis

DNA was visualised on an agarose gel containing 1.0 - 2.5% (w/v) agarose (Lonza, Switzerland) dissolved in TAE buffer (40mM Tris-acetate, pH 8.0, and 1mM EDTA) with 0.004% ethidium bromide. DNA samples were prepared in x6 blue/orange loading buffer (Promega, U.K.). To determine the size of the DNA bands, 1Kb and 100bp DNA markers (Promega, Southampton, U.K.) were run alongside any experimental samples. The agarose gels were electrophoresed in TAE buffer at 80-100V until DNA were separated; the DNA bands were then visualised using a UV transilluminator. The required DNA bands were excised and purified using the Wizard® SV Gel and PCR DNA Purification System (Promega) following the supplier's instructions.

### 2.2.3 A-tailing of the blunt-ended PCR products

The PCR products generated by Phusion polymerase are blunt-ended; therefore, they require the addition of a 3' adenine before being ligated into pGem-T. Taq polymerase (Genosys Ltd, U.K.) was used to A-tail the 3' end of the DNA following the instructions detailed in the Promega pGem-T manual; each A-tailing reaction contained 7µl DNA, 1x standard Taq buffer, 0.2mM dATPs, and 0.5U Taq, this mixture was incubated at 70°C for 30 min. The A-tailed DNA was then purified using the Wizard® SV Gel and PCR DNA Purification System (Promega) and ligated into pGem-T Easy.

### 2.2.4 DNA plasmid ligation

The A-tailed PCR products and restriction-digested gene fragments (Chapter 2.2.5) were ligated into either pGem-T or pET vectors, respectively. A typical 10µl ligation reaction contained 0.5-1.5U T4 DNA ligase, 1x ligase buffer (both Promega), 7.5µl purified DNA fragment, and 1µl of either pGEM-T Easy (Promega) or digested pET vectors (Novagen, Merck, U.K.). This mixture was incubated overnight at 16°C, along with a control ligation containing vector with no insert DNA.

### 2.2.5 Restriction digest

All restriction enzymes and buffers used were supplied by Promega or New England Biolabs (Ipswich, U.K.). A typical restriction digest contained 0.25U enzyme, 2µl recommended buffer, 5-10µl DNA, and $dH_2O$ up to 20µl. The reaction mixture was then incubated at 37°C for 2-4 h. In the case of pET vector digestion, to prevent the vector undergoing self-ligation, the 5'phosphate was removed by the addition of 0.5U Shrimp alkaline phosphatase (Roche, U.K.) during the final hour of the digestion.

### 2.2.6 Transformation of electrocompetent and chemically-competent *E. coli* cells

The plasmid DNA was ethanol precipitated, to remove salts and concentrate the DNA, before being transformed into the required *E. coli* cells. A typical ethanol precipitation mixture contained 10µl ligation mixture (Chapter 2.2.3), 66µl absolute ethanol, 0.1M sodium acetate, pH 5.2, and 1µl dextran blue. The mixture was incubated at -20°C for 30 min and then centrifuged at 13400 x g at 4°C to pellet the DNA. The precipitated DNA was washed with 100µl of 70% ethanol and re-pelleted before being left to dry. The plasmid was then re-suspended in 5µl Milli-Q water.

#### *2.2.6.1 Transformation of electrocompetent cells*

0.5µl DNA was added to 40µl of electrocompetent cells and incubated on ice for 5 min. These cells were then transferred into a 10mm pathlength Genepuler™ electroporation cuvette (BioRad, U.K.) before being electroporated using the Micropulser™ electroporator set to Ec1, the *E. coli* setting of approximately 1.8kV for 3-5 milliseconds. The cells were incubated for 1 h in 1ml SOC medium (LB medium supplemented with 20 mM glucose, 2.5 mM KCl, 20 mM $Mg^{2+}$ [from stock: 1M $MgCl_2.6H_2O$, 1M $MgSO_4.7H_2O$]) at 37°C. The cells containing the pET vectors were spread onto LB agar plates supplemented with the

required antibiotic. For blue/white selection of the JM109 cells containing the pGEM-T easy vector, the LB agar plates supplemented with 100µg/ml carbenicillin, with 100µl 100mM IPTG and 20µl 50mg/ml X-Gal was spread on the surface.

#### *2.2.6.2 Transformation of chemically-competent cells*

The *E. coli* BL21 and ArcticExpress™ expression strains were chemically transformed with the required pET vector. Approximately 50 ng (1-5 µl) of ethanol-precipitated DNA was added to 20µl of cells; this mixture was then incubated on ice for 30 min. The cells were incubated at 42°C for 45 sec, and then immediately cooled on ice. The cells were resuspended in 1ml SOC medium and incubated at 37°C for 1 h; the cells were then spread onto antibiotic-supplemented LB agar plates.

### 2.2.7 Colony PCR screens

Colony PCR reactions were performed when a large number of colonies were produced from transformations of newly-ligated pGEM-T easy or pET vector plasmids. A typical reaction contained 10µl of Taq Master Mix (GeneSys Ltd, Surrey, U.K.) [Taq polymerase, Taq buffer, dNTPs and loading dye] diluted with 9µl water and 0.5µl of the forward and reverse gene specific primers. The genomic DNA was introduced to the PCR mixture by the addition of a few cells from each colony; the cells were lysed and DNA released during the initial denaturation step shown in Table 2.3. The PCR products were analysed by agarose gel electrophoresis, and any colonies that resulted in a PCR product of the expected gene size were subjected to a double restriction digest as described in Chapter 2.2.5.

| | | Temperature (°C) | Time |
|---|---|---|---|
| | **Initial denaturation** | 95 | 30 sec |
| X40 cycles | **Denaturation** | 95 | 30 sec |
| | **Annealing** | 55 | 30 sec |
| | **Elongation** | 68 | X min |
| | **Final elongation** | 68 | 10 min |

**Table 2.3 The temperatures and times of each stage of Taq polymerase PCR**. The elongation time (X) is dependent on the size of the gene (1 min per 1Kb).

### 2.2.8 DNA sequencing

The genes inserted into the multiple cloning regions of both pGem-T Easy and pET vectors were sequenced, in the forward and reverse directions, using commercially available primers. Inserts in pGem-T Easy were sequenced using the T7F and SP6 primers, and all pET vector inserts were sequenced using the T7F and T7R primers. Internal gene-specific primers were designed for genes longer then 1.5Mbp. The sequencing reactions were performed by SourceBioscience plc (Oxford, U.K.).

## 2.3 Protein expression and purification

### 2.3.1 Protein expression

The pET vectors containing the genes of interest were transformed into *E. coli* Rosetta(DE3)pLysS, *E. coli* BL21 or Arctic Express™ for protein expression. A 10ml LB medium starter culture, supplemented with the required antibiotics, was inoculated with a single colony and grown overnight; 2ml of this culture were used to inoculate 500ml of LB medium, in 2L flasks, supplemented with the required antibiotic. All cultures were grown at 37°C until they reached an $OD_{600}$ ≈0.6-1.0. The conditions after this point varied depending on the protein being expressed and the *E. coli* expression strain. The cultures containing the *Tp. acidophilum* genes were grown in conditions described in Table 2.4 for approximately 20 h, before the cells were harvested by centrifugation at 13400xg, 4°C for 15 min.

| Protein | Induction method | Post induction temperature (°C) |
|---|---|---|
| E1 | 1mM IPTG | 25 |
| E2 | 1mM IPTG | 25 |
| E2cat | 1mM IPTG | 37 |
| E3 | Not induced (Aass, 2003) | 37 |

**Table 2.4 The conditions used for the heterologous expressions of the *Tp. acidophilum* complex proteins in *E. coli*.** The mutated versions of the E2 and E2 catalytic domain (Chapters 7 and 8) were expressed in the same conditions as the non-mutated protein.

### 2.3.2 Preparation of soluble cell extract

The pelleted cells were resuspended in the required buffer (0.25-0.50 % (v/v) of the initial culture); E1, E2 and E3 samples were suspended in his-bind buffer (300mM NaCl, 50mM Tris-HCL, 20mM imidazole (Acros Organics, U.K.), pH8.0), and the E2cat samples in 50mM

Tris-HCl, pH8.8. Complete protease inhibitor cocktail, EDTA-free, tablets (Roche) were added to all samples and 1mM EDTA was also added to the E2cat sample. The cells were lysed by sonication, whilst remaining on ice, to release the cytoplasmic proteins and shear the genomic DNA.

#### *2.3.2.1 Heat treatment*

Heat treatments were used immediately after lysing the cells, to remove a large proportion of the contaminating *E. coli* proteins. The *Tp. acidophilum* protein samples were incubated at 55°C for 10 min. The cell lysate was then centrifuged at 13400 x g at 4°C for 15-20 min. The clear supernatant was then passed through a 0.45µm filter (Millipore, U.K.).

#### *2.3.2.2 Nickel-affinity chromatography*

Nickel-affinity chromatography was used to purify the N-terminal histidine-tagged E1, E2, and E3 proteins. A PolyPrep chromatography column (BioRad, U.K.) was filled with 1ml chelating cellulose (Bioline, U.K.) and washed with 5ml of Milli-Q water, followed by 2ml 0.4M $NiSO_4$; the excess $NiSO_4$ was removed with Milli-Q water and the column was finally equilibrated with his-bind buffer (50mM Tris-HCl, pH 8.0, 20mM imidazole (Acros organics, U.K.), and 300mM NaCl). The filtered supernatant was loaded onto the column and allowed to flow through; the solution that passed through the column was collected and loaded again to ensure all his-tagged proteins had bound. The unbound proteins that remained in the column were eluted with 5ml of his-bind buffer. The bound proteins were then removed by increasing the concentration of imidazole in the elution buffer; the his-elute buffer (50mM Tris-HCl, pH 8.0, 1M imidazole, and 300mM NaCl,) and the his-bind buffer were mixed together to produce an imidazole concentration gradient, usually 5% and 10% (v/v) his-elute to his-bind solutions. The undiluted 100% his-elute buffer was added and the eluent collected from the column in five 1ml aliquots.

#### *2.3.2.3 Analytical Gel filtration*

The Superdex 200, 10-300GL (GE Healthcare) was used to purify and determine the size of the recombinant proteins. A maximum of 0.5ml of protein solution was loaded on to the Superdex 200 column, which had been equilibrated with the required buffer. Gel filtration was used as the final E2cat protein purification step (Chapter 4); in this case the column was equilibrated in 50mM Tris-HCl, pH 8.8, and 100mM NaCl. The molecular weights of the eluted protein peaks were determined by comparison with standard proteins of known molecular weight: Carbonic Anhydrase 29 kDa, Bovine Albumin 66 kDa,

Alcohol Dehydrogenase 150 kDa, β-Amylase 200 kDa, Apoferritin 443 kDa, and Thyroglobulin 669 kDa. Dextran Blue (2000kDa; Sigma-Aldrich) was used to measure the void volume of the column.

### 2.3.3 SDS-PAGE

Proteins were visualised by SDS-PAGE using 10% or 12.5% acrylamide gels.

10% Separating gel: 4ml Milli-Q water, 3.3ml 30% (v/v) acrylamide (BioRad), 2.5ml 1.5M Tris-HCl pH8.8, 100µl 10% SDS, 100µl 10% (w/v) ammonium persulphate and 10µl tetramethylethylenediamine (TEMED) (National diagnostics).

Stacking gel - 1.5ml Milli-Q water, 0.3ml 30% (v/v) acrylamide , 0.5ml 0.5M Tris-HCl pH6.8, 20µl 10% SDS, 20µl 10% (w/v) Ammonium persulphate, and 5µl TEMED.

The protein samples were heated at 95°C for 3 min with loading buffer (4x loading dye contained 4ml 0.5M Tris-HCl, pH 6.8, 2ml 10% SDS, 4ml glycerol, 100µl β-mercapoethanol (Sigma-Aldrich), 8µg bromophenol blue); each sample was briefly centrifuged at 13400 x g before being loaded onto a gel. The size of the protein bands were determined by comparison with the BioRad (U.K.) broad-range standard molecular markers: 200 kDa myosin, 116 kDa β-galactosidase, 97 kDa phosphorylase b, 66kDa serum albumin, 45 kDa ovalbumin, 31 kDa carbonic anhydrase, 21.5 kDa trypsin inhibitor, 14.4 kDa lysozyme, and 7 kDa apoferitin.

The electrophoresis tank was filled with Tris/glycine buffer containing 0.02M Tris-HCl, 0.192M glycine, and 0.1% (w/v) SDS. The gels were run at 200V until the marker dye had reached the bottom of the gel. The gels were stained with Coomassie stain (50% (v/v) Methanol, 10% (v/v) Acetic acid, and 0.01% (w/v) Coomassie Brilliant Blue R 250 (Sigma-Aldrich)), and then the excess stain was removed from the gel by incubation in destain solution containing 30% (v/v) Methanol and 10% (v/v) Acetic acid.

## 2.4 Determination of protein concentration

### 2.4.1 Bradford Assay

The Bradford assay was used to determine the protein concentration of any impure protein samples. A standard curve was produced by the addition of 200µl of Bio-Rad

reagent (Coomassie blue G 250) to 800µl of bovine serum albumin (Sigma-Aldrich) standards ranging between 2µg/ml and 10µg/ml. The absorbance of the standards and protein samples were measured at 595nm, in a Cary UV spectrophotometer.

### 2.4.2 Protein concentration determination by $A_{280}$

The concentration of pure protein samples were determined spectrophotometrically at 280nm on a Cary UV spectrophotometer, using the absorption coefficient predicted by Expasy ProtParam, based on the amino acid sequence (Gasteiger et al., 2003). Non-specific light scattering was taken into account when measuring the E2 protein concentration. That is, the apparent absorbance was measured at 280, 320 and 360nm, where that observed at 360 and 320nm is due to scattering of the beam rather than absorbance. A linear extrapolation from 360-320 to 280nm then allowed a correction to be made to give the true $A_{280nm}$. In the case of E3 enzyme, the FAD concentration was determined by the absorbance at 455nm using the molar absorption coefficient 11.3 $x10^3$ $M^{-1}cm^{-1}$.

## 2.5 Enzyme assays

Unless otherwise stated, all material was kept on ice, all buffers were pre-incubated to the required temperature before each experiment, and assays were carried out using a Cary 300 Bio or Cary 50 Bio spectrophotometer. The total volume of all assays was 1ml.

### 2.5.1 2-Oxoacid decarboxylase (E1) spectrophotometric assay

The *Tp. acidophilum* E1 enzyme was first incubated at 55°C for 10 min in preheated E1 buffer (20mM potassium phosphate, pH 7.0, 2mM $MgCl_2$, and 0.2mM thiamine pyrophosphate). The background rate was determined after the addition of 1µM 2,6-dichlorophenolindophenol (DCPIP) (Sigma-Aldrich), an artificial electron acceptor that allows E1 activity in the absence of the E2 attached lipoic acid moiety. The reaction was started by the addition of 0.4mM 2-oxoacid (Sigma-Aldrich). The rate of DCPIP reduction was measured by the decrease in the absorbance at 595nm, using the absorption coefficient of 2.2$x10^3$ $M^{-1}cm^{-1}$.

### 2.5.2 Dihydrolipoamide S-acetyltransferase (E2) spectrophotometric coupled assay

E2 enzyme activity was measured by coupling the E2-transacylase activity with a phosphotransacetylase (PTA) from *Geobacillus stearothermophilus* (EC Number: 2.3.1.8) (Sigma-Aldrich), based on the assay described by Packman et al. (1984). The production of thioester bonds were measured by the absorbance change at 233nm, using the absorption coefficient of $5.4 \times 10^3$ $M^{-1}cm^{-1}$ (Srere and Kosicki, 1961). Each 1ml assay contained 0.1M Tris-HCl, pH 7.6, 4.0mM DL-dihydrolipoamide (DHlip), 0.1mM coenzyme A, 10mM acetyl phosphate, and 7.5U phosphotransacetylase; this reaction was allowed to reach equilibrium at 55°C before the E2 enzyme was added. The E2 enzyme then converted the acetyl-CoA to acetyl-dihydrolipoamide, with the excess PTA replacing any acetyl-CoA used, resulting in an overall increase in total concentration of thioester bonds (an example assay is shown in Figure 2.1) (Packman et al., 1984).

First reaction - catalysed by the phosphotransacetylase (PTA):

CoA + acetyl phosphate ↔ **acetyl-CoA** + Pi

Second reaction - catalysed by E2 dihydrolipoamide transacetylase:

**acetyl-CoA** + DHlip ↔ Acetyl-dihydrolipoamide + CoA

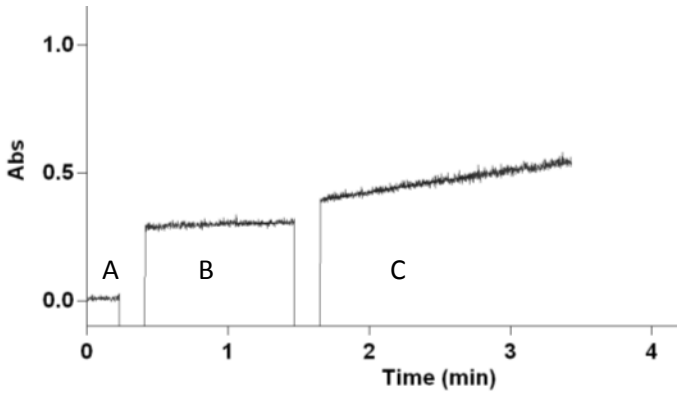



**Figure 2.1 An example of the coupled E2-PTA assay 233nm trace.** Rate A is the background reaction with all substrates but no enzymes; rate B is as a result of the addition of excess PTA, resulting in an equilibrium of CoA + acetyl phosphate and acetyl-CoA and Pi. Rate C is due to the addition of E2 enzyme which converts acetyl-CoA into Acetyl-dihydrolipoamide.

### 2.5.3 Dihydrolipoamide dehydrogenase (E3) spectrophotometric assay

The rate DHlip oxidation was measured by the reduction of $NAD^+$ to NADH at 340nm, using the absorption coefficient of $6.22x10^3$ $M^{-1}cm^{-1}$. The E3 assay buffer (50mM sodium phosphate, pH7.0) was preheated to 55°C. The background rate was measured after the addition of 0.4mM $NAD^+$ and 0.4mM dihydrolipoamide (DHlip) to the cuvette, after which the reaction was started with the addition of E3 enzyme.

#### *2.5.3.1 Preparation of dihydrolipoamide*

DL-Lipoamide was reduced by $NaBH_4$ to DL-dihydrolipoamide (DHlip) following the method described by Reed *et al*. (1958). The DL-lipoamide (800mg) was dissolved in 40ml methanol:water (10:1) and mixed with 800mg $NaBH_4$ dissolved in 4ml cold $dH_2O$. The solution was the stirred on ice for 2 h or until colour changed from yellow to white. The solution was then acidified by the addition of 1M HCl until the effervescence stopped. The product was extracted three times with 50ml chloroform; the chloroform was removed from the dihydrolipoamide using a rotary evaporator. The resulting crystals were dissolved in 40ml toluene and gentle warmed. 16ml hexane was added to this solution to precipitate dihydrolipoamide; the precipitated DHlip was collected and dried by vacuum filtration.

#### *2.5.3.2 Determination of purity*

The purity of the DHlip was determined spectrophotometrically by reacting with 0.2mM 5',5'-dithiobis-2-nitrobenzoate (DTNB) with 0.05mM dihydrolipoamide (0.1mM thiol groups) in 1ml20mM Tris-HCl, pH 8.0. A 100% pure sample of DHlip would be expected to have an absorbance at 412nm of 1.36 ($E_m$ = $13.6x10^3$ $M^{-1}$ $cm^{-1}$) (Reed et al., 1958).

### 2.5.4 Branched chain 2-oxoacid dehydrogenase complex assay

The whole OADHC assay was performed in buffer containing 50mM potassium phosphate, pH 7.0, 2.5mM $NAD^+$, 1mM $MgCl_2$, and 0.2mM TPP. The $E1\alpha_2\beta_2$, E2 and $E3\alpha_2$ enzymes were added to the preheated buffer in the required molar ratio and incubated for 10 min at 55°C to allow the complex to assemble (Heath, 2006). The reaction was started by the addition of 0.13mM CoA and the 2-oxoacid substrate, e.g. 2mM 3-methyl-2-oxopentanoic acid. The rate of conversion of $NAD^+$ to $NADH+H^+$ was measured at 340nm, using the absorption coefficient of $6.22x10^3$ $M^{-1}cm^{-1}$ (Reed et al., 1958).

### 2.5.5 Determination of Michaelis-Menten kinetic parameters

The Michaelis-Menten kinetic parameters, Michaelis constant ($K_m$) and maximum velocity ($V_{max}$), were determined by assaying the enzymes in saturating conditions for all but one substrate, which was varied each time. The specific activity (U/mg) was calculated for each substrate concentration, where one Unit of enzyme activity (U) is defined as 1 µmol of substrate disappearance or product appearance per minute. The $K_m$ and $V_{max}$ values were calculated by SigmaPlot 12 (Systat Software Inc) by fitting the data to the Michaelis-Menten equation and the Hanes-Woolf derivative, shown below:

Michaelis-Menten equation:

$$v = \frac{V_{\max} \times [\mathrm{S}]}{K_m + [\mathrm{S}]}$$

Hanes-Woolf equation:

$$\frac{[\mathrm{S}]}{v} = \frac{1}{V_{\max}}[\mathrm{S}] + \frac{K_m}{V_{\max}}$$

Where $v$ is the velocity of the reaction at a given substrate concentration [S], $K_m$ is the Michaelis constant and $V_{max}$ is the maximum velocity.

# Chapter 3: Characterisation of the *Tp. acidophilum* 2-oxoacid dehydrogenase complex

## 3.1 Introduction

The *Tp. acidophilum* branched-chain 2-oxoacid dehydrogenase complex is composed of three enzymes: E1 (2-oxoacid decarboxylase), E2 (dihydrolipoamide transacetylase), and E3 (dihydrolipoamide dehydrogenase); consequently, the thermostability and thermoactivity of the assembled complex is dependent upon the stability of these individual subunits and the interactions between them. This chapter studied the thermostability and thermoactivity of the individual recombinant enzymes and the assembled complex. The *in vitro* study of the assembly of the *Tp. acidophilum* OADHC with varying stoichiometries of E1 and E3 compared to E2 subunits, and the effect on the relative activity of the complex was also investigated. E3 activity has been detected in *Tp. acidophilum* cell lysates and the complex genes have been shown to be transcribed and translated, and E1 and E2 appear to assemble into a large complex (Sun et al., 2007, Sun et al., 2011). However, OADHC activity has never been detected in any archaeal cell lysate, and therefore the aim of this chapter was to understand the thermostability, thermoactivity and assembly of this large multienzyme complex in an attempt to explain the lack of detectable activity.

## 3.2 Methods

### 3.2.1 Cloning of E1β

The E1β gene was released from the pGem-T Easy plasmid by digestion with *Nhe* I and *Bam* HI, and was then ligated into pET24a, resulting in a gene encoding the recombinant E1β enzyme without any additional tags. The E1β-pET24a and E1α-pET19b plasmids were then co-transformed into *E. coli* Rosetta (DE3) pLysS for co-expression.

### 3.2.2 Expression of E1, E2 and E3

The individual enzymes were all expressed as described in Chapter 2.3.1 with the exception of the E2; when lipoylated protein was required in this case, the medium was supplemented with 0.2mM DL-lipoic acid.

### 3.2.3 Mass spectrometric analysis

Purified protein sample was buffer exchanged by dilution into 50ml 20mM acetic acid and then concentrated to 1ml followed by a second dilution into another 50ml of 20mM acetic acid. The sample was then concentrated to less than 500µl, giving an E2 sample with a final concentration of 20 mg/ml. The samples were then analysed by on a MicrOTOF$^{TM}$, a time-of-flight mass spectrometer coupled with an electrospray source (ESI-TOF) (Bruker Daltonics Bremen, Germany). Samples were injected into the machine via a syringe pump and data acquired for an average of 5 min. The machine was calibrated after the samples were analysed using a sodium formate solution. The results were analysed using the DataAnalysis program (Bruker Daltonics). All samples were loaded and analysed by Dr A. Lubben (Department of Chemistry, University of Bath).

### 3.2.4 Western blot analysis

Western blot analysis was used to determine if the E2 protein had been lipoylated *in vivo* during recombinant expression. Samples were first run on a 10% SDS-PAGE gel using loading dye without β-mercaptoethanol to ensure the lipoic acid was not in a reduced form. The protein bands were then transferred to an Immobilon-P membrane (Millipore) by wet transfer in 20% (v/v) methanol, 192mM glycine, and 25mM Tris-HCl using the Mini Blot system (BioRad) at 100V for 1 h, at 4°C.

The membrane was then removed and incubated over night on a platform shaker (Stuart Scientific, U.K.), at 4°C, in phosphate buffered saline (PBS) solution pH 7.3 (160mM NaCl, 3mM KCl, 8mM $Na_2HPO_4$, and 1mM $KH_2PO_4$)(Oxoid) with 0.1% Tween 20, and 5%(v/v) milk powder to block any nonspecific antibody binding sites. The membrane was washed with PBS wash solution (PBS with 0.2% Tween 20) before incubation for 1 h with the rabbit anti-lipoic antibody (Calbiochem), which was diluted 5000 fold in blocking solution. The membrane was washed with PBS wash solution and incubated with the secondary antibody, anti-rabbit IgG–peroxidise antibody (diluted the same as the primary antibody). After a final PBS wash, the membrane was incubated with Immobilon Western Chemiluminescent horseradish peroxidase (HRP) substrate at room temperature for 2 min. The western blot was then developed by exposing the membrane to X-ray film (Fujitsu) for between 5 s and 1 min.

### 3.2.5 Heat precipitation assays

Thermal precipitation experiments were carried out in a Mastercycler PCR machine (Eppendorf) using thin-walled 0.5ml Eppendorf tubes. Protein samples were heated at various temperatures for a set time, usually 10 min; the soluble and insoluble proteins were then separated by centrifugation at 13 400 x g at 4°C for 10 min. The insoluble protein was then re-suspended in an equal volume of 8M urea; both the soluble and insoluble fractions for each temperature were analysed by SDS-PAGE.

### 3.2.6 Thermal inactivation assays

The thermal inactivation assays were conducted in a Mastercycler PCR machine (Eppendorf) using thin-walled 0.5ml Eppendorf tubes. The protein was heated in the PCR machine at various temperatures, the lid set to 105°C to prevent evaporation. Sample tubes were removed periodically and cooled on ice. Once fully cooled the tubes were centrifuged at 16000 x g for 5 min, 4°C. The enzymes were assayed (see Chapter 2.5 for assay details), to determine the percentage activity of the heat-treated protein compared to the activity of the protein which had not been heated. The E3 thermal inactivation experiments were carried in the presence and absence of 0.4 mM DHlipoamide.

### 3.2.7 Temperature optimum

The E1 and OADHC enzymes were pre-heated in the E1 or complex assay buffer: containing phosphate buffer, TPP and $MgCl_2$, respectively, for 10 min at 55°C prior to being assayed. The reactions were started with the addition of enzyme rather than substrate as previously described for the E1 and complex assay and the temperature was recorded at the start and end of each assay.

### 3.2.8 Determination of OADHC component concentrations and stoichiometry assays

The concentrations of the purified enzymes were determined by measuring the absorbance at 280nm. The $A_{280}$ was then converted to a molar concentration of the active protein unit ($E1\alpha_2\beta_2$, $E2\alpha$, and $E3\alpha_2$) using the molar absorbance coefficients predicted from the sequence (Gasteiger et al., 2003).

The required amounts of $E1\alpha_2\beta_2$, $E2\alpha$, and $E3\alpha_2$ were mixed in the cuvette containing complex assay buffer and incubated for 10 min at 55°C; whole complex 2-oxoacid dehydrogenase activity was assayed by measuring the conversion of $NAD^+$ to NADH, after the addition of 0.13mM CoA and 2mM 3-methyl-2-oxopentaoic acid (Chapter 2.5.4).

### 3.2.9 Gel filtration

The required ratio of $E1\alpha_2\beta_2$, $E2\alpha$, and $E3\alpha_2$ were mixed and diluted 1:1 in 50mM potassium phosphate, pH 7.7, 1mM $MgCl_2$, and 2mM 3-methyl-2-oxopentanoic acid. The complex mixture was then heated at 55°C for 10 min to allow assembly and cofactor binding; any heat precipitated proteins were removed by centrifugation at 16 000 x g and filtration through a 0.45µm filter. The soluble OADHC sample was then loaded onto a Superdex 200 column equilibrated with 50mM potassium phosphate, pH 7.7, 1mM $MgCl_2$, and 2mM 3-methyl-2-oxopentanoic acid; 0.25-0.5ml fractions were collected. The E1 and OADHC assays were modified for this experiment due to the presence of 3-methyl-2-oxopentaonic acid in the gel filtration buffer: the E1 assay was started by the addition of DCPIP and the OADHC assays by the addition of CoA.

## 3.3 Results and Discussion

### 3.3.1. E1 2-oxoacid decarboxylase

#### *3.3.1.1 Cloning of Thermoplasma acidophilum E1β into pET24a*

The E1α and E1β genes were cloned by Dr. Heath into pET19b and pET28a, respectively (Heath et al., 2004), resulting in 4 N-terminal his-tags per $E1\alpha_2\beta_2$ enzyme. To reduce the number of his-tags in the assembled $E1\alpha_2\beta_2$ protein, the E1β gene was cloned in to pET24a (Figure 3.1) using *Nde* I and *Bam* HI, instead of pET28a. The pET19b (carbenicillin resistant) and pET24a (kanamycin resistant) vectors containing the his tagged E1α and non-tagged E1β genes were then co-transformed into *E. coli* Rosetta (DE3) pLysS cells for co-expression.

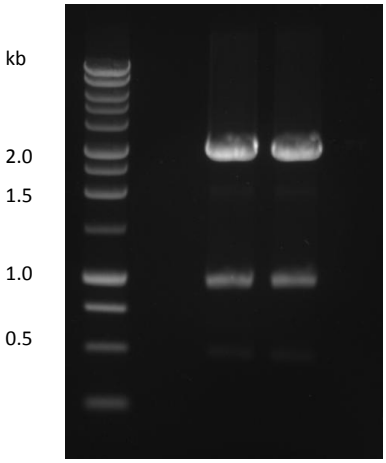



**Figure 3.1 Agarose gel of pGem-T cut with *Nhe* I and *Bam* HI to release the E1β gene (960bp)**

### *3.3.1.2 Expression and purification of E1α$_2$β$_2$*

The untagged E1β protein was co-expressed with the E1α protein in *E. coli* Rosetta(DE3)pLysS. Soluble E1α and E1β proteins were co-purified by nickel-affinity chromatography using the E1α N-terminal his-tag (Figure 3.2).

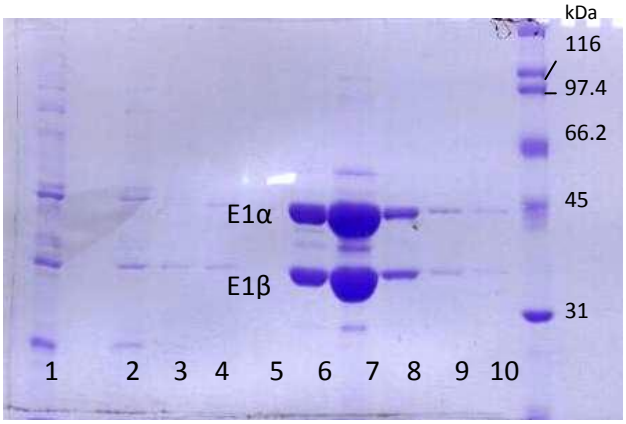



**Figure 3.2 SDS-PAGE of the co-purification of the E1α and β polypeptides.** Lane 1: flow through; lane 2: the his-bind buffer wash; lanes 3 and 4: the 5% and 10% his-elute washes; lanes 5 to 9: the 1ml fractions eluted with 100% his-elute buffer: lane 10 standard protein markers, with their $M_w$ values.

### *3.3.1.3 Characterisation of E1 2-oxoacid decarboxylase*

The removal of the his-tag from the E1β subunit may have altered the kinetic parameters of the E1 enzyme; therefore, the $K_m$ and $V_{max}$ values were determined for several 2-oxoacid substrates except 2-oxoglutarate with which the enzyme demonstrated no activity (Table 3.1). The assays were conducted as described in Chapter 2.5.1 with the exception of the 2-oxoacid substrate concentration which was varied each time. The activity data were analysed using SigmaPlot 12 (Systat Software Inc) shown in figures 3.3-3.8.

| **Substrate** | **$K_m$ (μM)** | **$V_{max}$ (U/mg)** |
|---|---|---|
| **3-methyl-2-oxopentanoic acid** | 88.8 ± 3.3 | 0.881 ± 0.010 |
| **4-methyl-2-oxopentanoic acid** | 36.2 ± 1.2 | 0.345 ± 0.003 |
| **Pyruvate** | 1732.3 ±31.9 | 0.249 ± 0.001 |

**Table 3.1 Kinetic parameters of the E1α$_2$β$_2$ enzyme.** 1 Unit of enzyme activity is defined as 1μmol of DCPIP reduced per min.

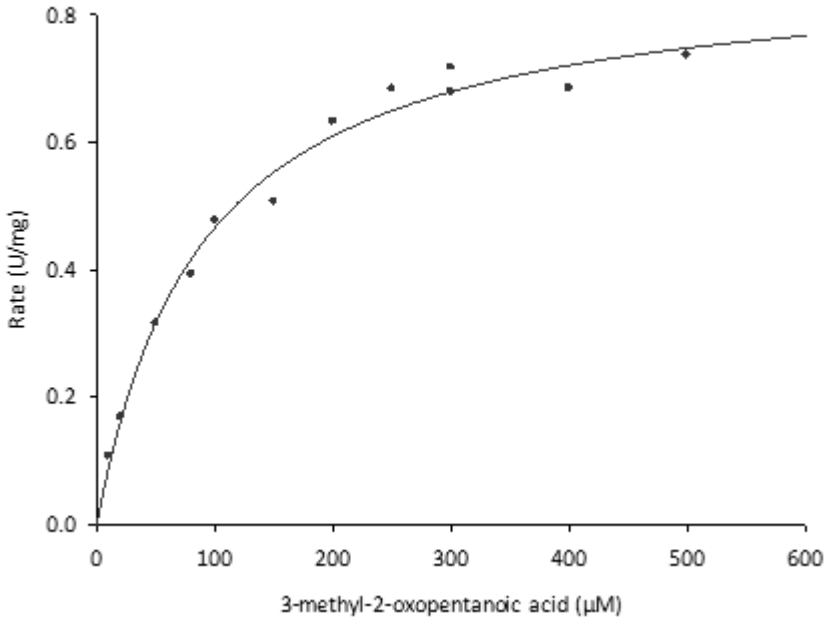



**Figure 3.3 Michaelis-Menten plot of E1 activity with 3-methyl-2-oxopentanoic acid**

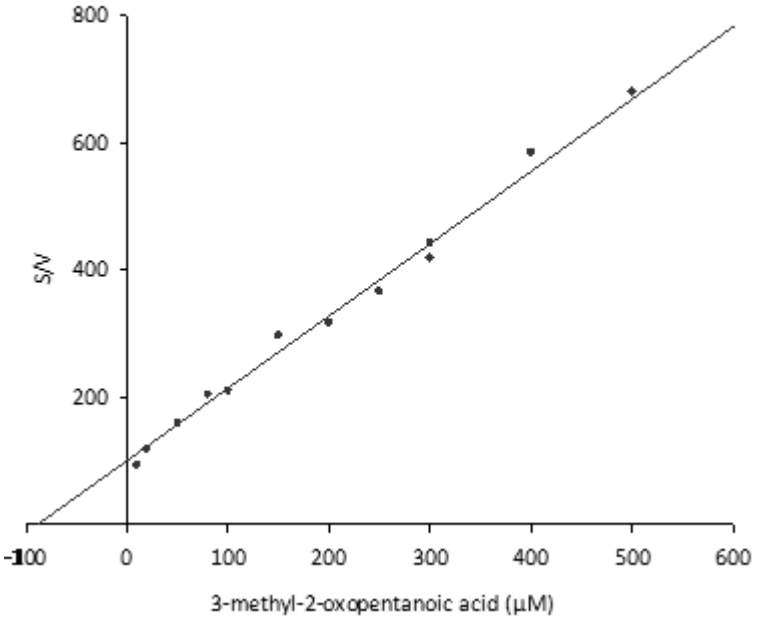



**Figure 3.4 Hanes-Woolf plot of E1 activity with 3-methyl-2-oxopentanoic acid**

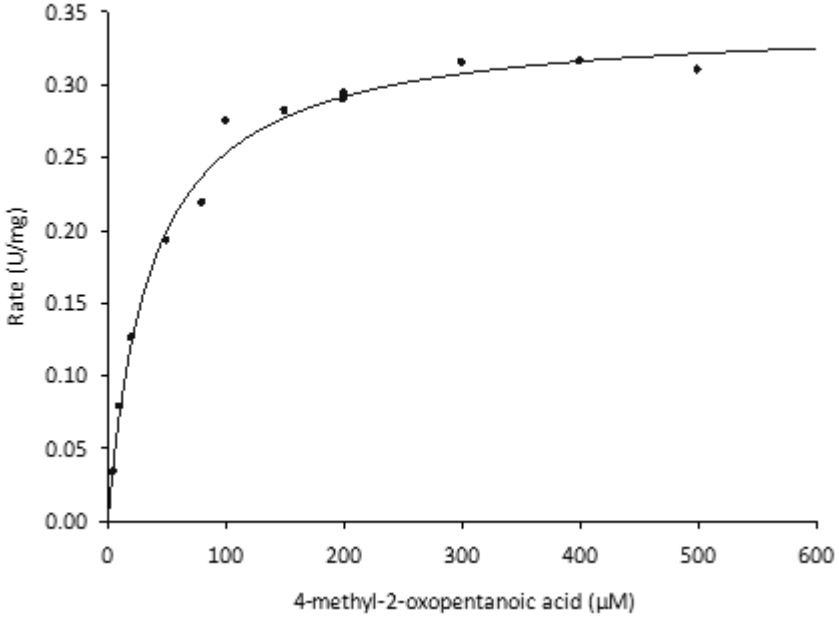



**Figure 3.5 Michaelis-Menten plot of E1 activity with 4-methyl-2-oxopentanoic acid**

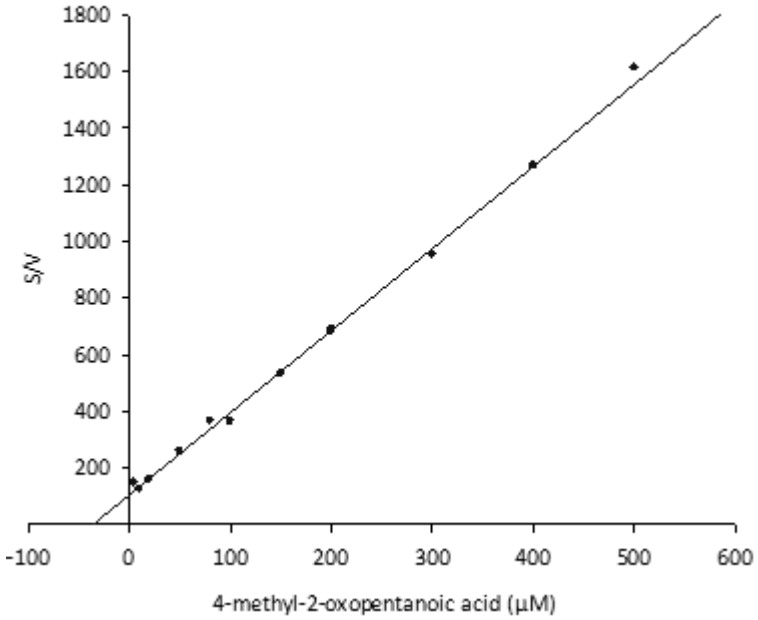



**Figure 3.6 Hanes-Woolf plot of E1 activity with 4-methyl-2-oxopentanoic acid**

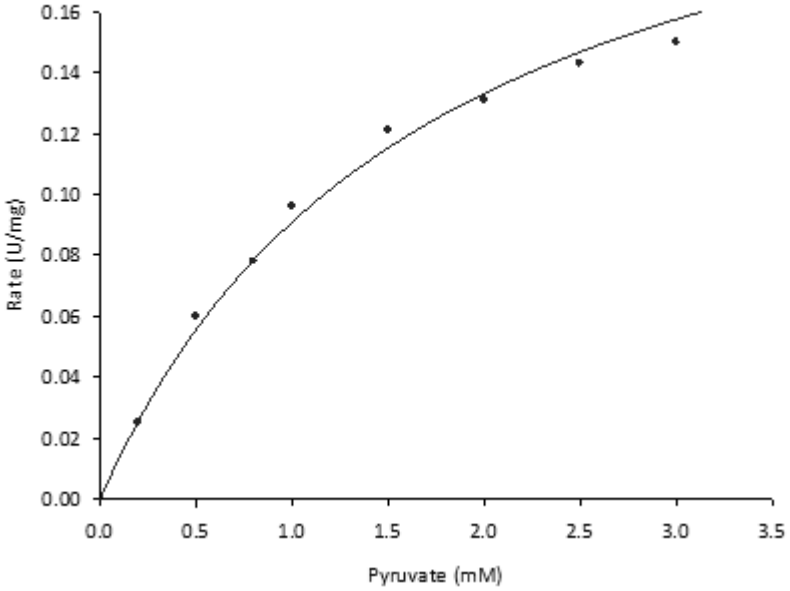



**Figure 3.7 Michaelis-Menten plot of E1 activity with pyruvate**

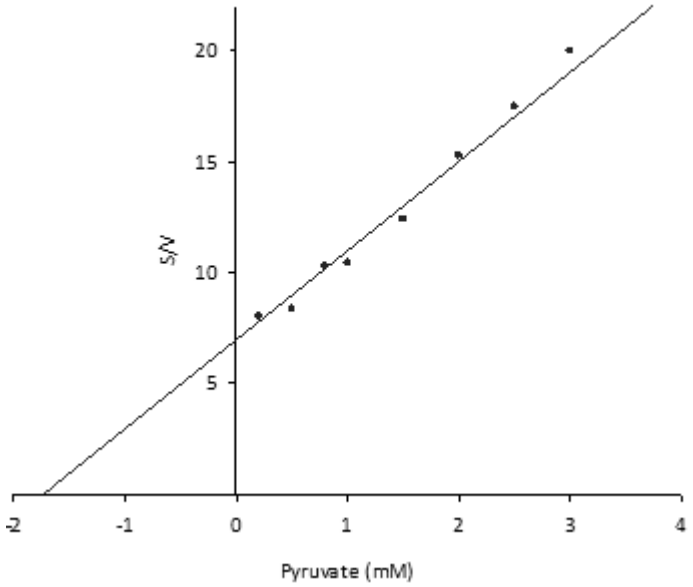



**Figure 3.8 Hanes-Woolf plot of the E1 activity obtained using pyruvate**

The E1β subunit, without a his-tag, co-purified with the E1α, resulting in a functional $E1\alpha_2\beta_2$ enzyme with half the number of his-tags of the previous recombinant E1. The order of substrate preference remained unchanged but in all cases the $V_{max}$ increased; e.g. the $V_{max}$ with 3-methyl-2-oxopentanoic acid increased from 0.38U/mg protein to 0.86U/mg protein. The $K_m$ of the branched chain 2-oxoacids decreased slightly, whereas, the $K_m$ for pyruvate increased from 596µM (Heath, 2006) to 1380µM.

#### *3.3.1.5 E1: Temperature optimum*

The temperature optimum of the recombinant *Tp. acidophilum* E1 enzyme was determined to be 72°C.

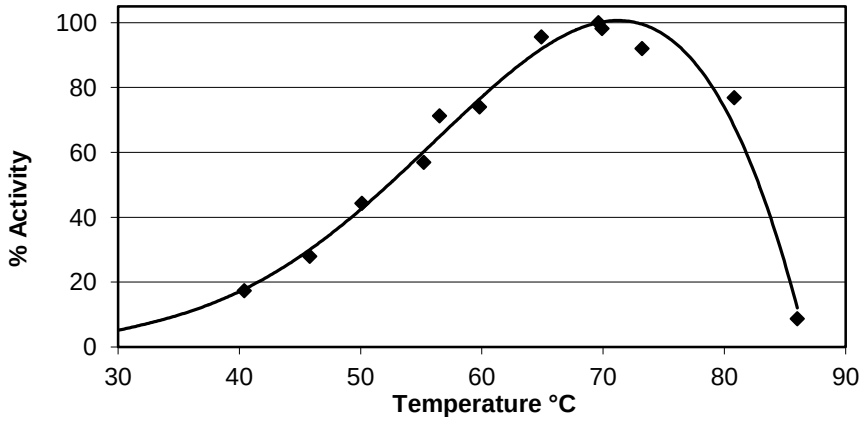



**Figure 3.9 Temperature optimum of the E1 enzyme.** Activity expressed as percentage of activity at 70°C

### *3.3.1.4 E1: Thermostability*

The thermostability of the E1 enzyme was measured by assaying heat-treated protein under optimal conditions. The reaction rate of enzyme that had not been heat-treated was taken to be 100%; the activity of the thermally-inactivated samples was compared to this value to determine the percentage activity of the sample after heating (Chapter 3.2.6). The E1 enzyme was diluted 3-fold in E1 assay buffer containing: 20mM potassium phosphate pH7.0, 2mM $MgCl_2$, and 0.2mM TPP.

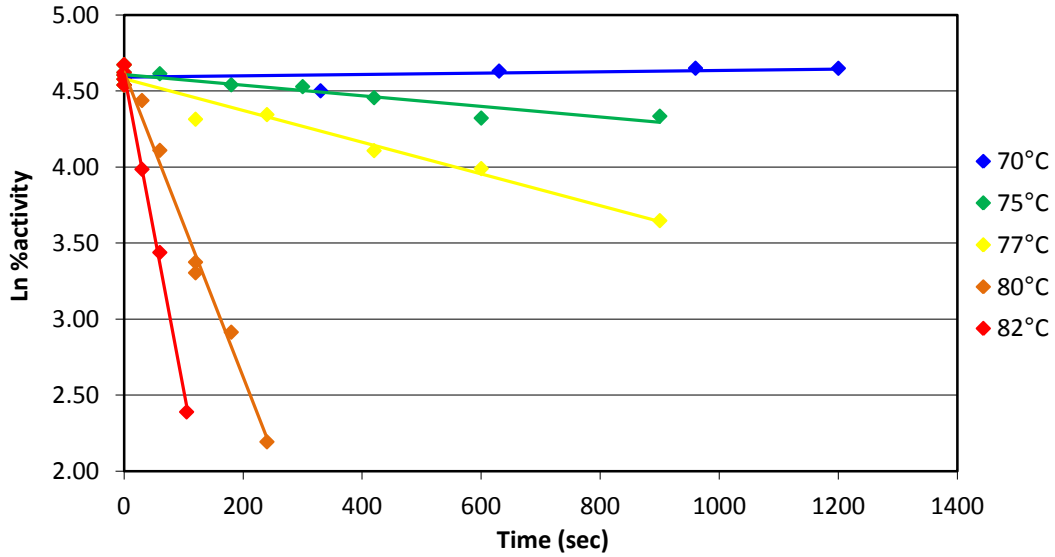



**Figure 3.10 Thermal inactivation plot of the E1 enzyme**. The E1 activity before heating is taken to be 100%. Samples were heated at a various temperature then cooled before the remaining activity was assayed under optimal conditions.

### 3.3.2 E2 Dihydrolipoamide transacetylase

#### *3.3.2.1 E2 expression, lipoylation and purification*

The E2 protein was purified to 85-95% purity by a heat-precipitation step followed by nickel-affinity chromatography, as shown by Figure 3.11.

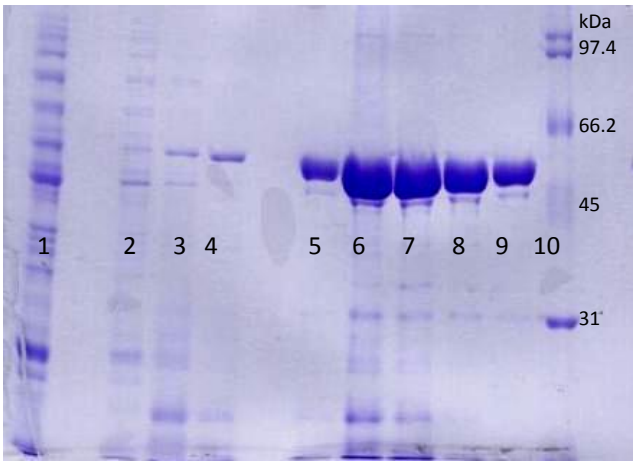



**Figure 3.11 SDS-PAGE of E2 purification by nickel affinity chromatography.** Lane 1: flow through protein that did not bind to the column; lanes 2 to 4: the his-bind buffer, 5% and 10% his-elute washes; lanes 5 to 9: his elute wash collected in 1ml fractions; lane 10: standard protein markers, with their molecular weight values.

The recombinant E2 protein must be lipoylated, by the *E. coli's* lipoylation system, to be catalytically active in the whole complex reaction. To increase the yield of lipoylated E2, an expression trial was carried out to determine if the amount of lipoylated recombinant E2 could be increased (Figures 3.12 and 3.13).

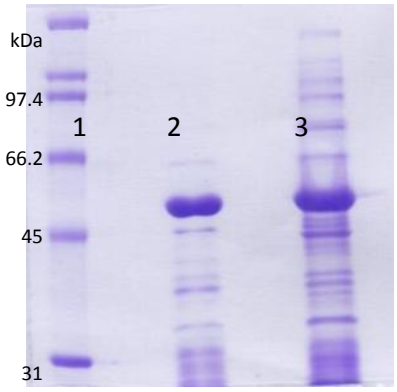



**Figure 3.12 SDS-PAGE of recombinantly expressed E2 protein supplemented with lipoic acid.** Lane 1: the standard protein markers. Lanes 2 and 3 contain the cell extract from medium supplemented with 0.2 and 0.4mM lipoic acid, respectively.


**Figure 3.13 Western blot of the recombinantly expressed E2 protein that has been lipoylated *in vivo*.** Lanes 1 and 2 contain the cell extract from medium supplemented with 0.2 and 0.4mM lipoic acid, respectively, stained by rabbit anti-lipoic acid and anti-rabbit antibodies.

Induction with 1mM IPTG increased the yield of lipoylated E2, this suggests that the *E. coli* lipoylation system is capable of dealing with the higher expression level. Two concentrations of supplemented lipoic acid (0.2 and 0.4mM) were also compared by western blot analysis (Figure 3.13); it was found that changing the concentration of lipoic acid in the medium did not appear to significantly affect the relative amount of recombinant E2 lipoylation.

To determine the percentage of recombinant E2 protein that has been lipoylated during expression, a purified sample was analysed by mass spectrometry. Figure 3.14 shows two major peaks at 46460Da and 46278Da, these two values correspond to the predicted molecular weight of the lipoylated (46464 Da) and non-lipoylated (46276Da) E2 protein without the start methionine.

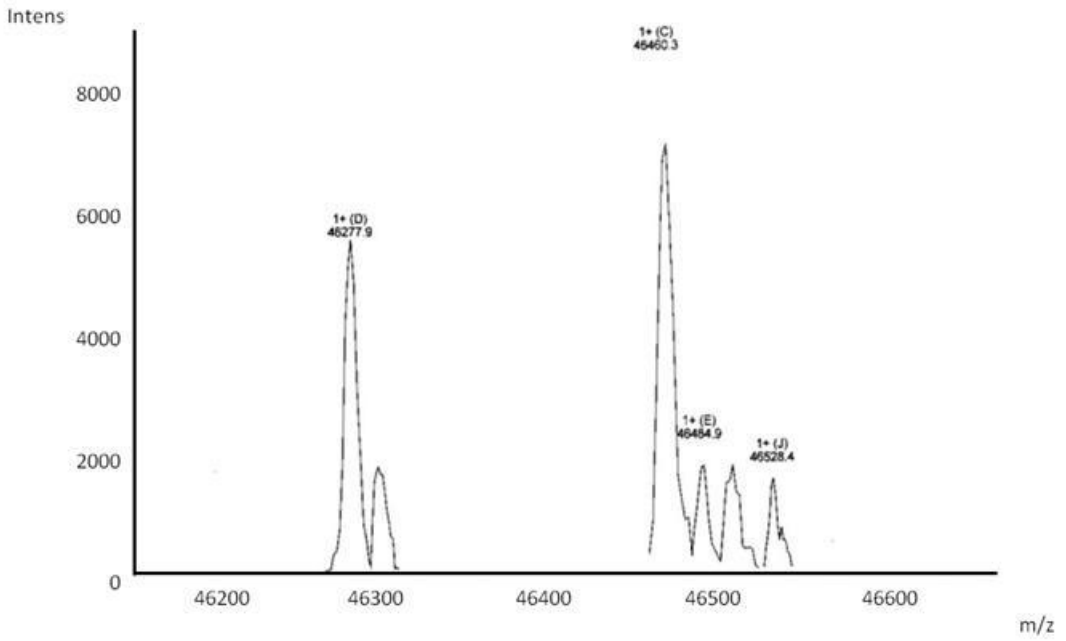



**Figure 3.14 Mass spectrometry analysis of recombinant E2 *in vivo* lipoylation.** The E2 protein without the start methionine is predicted to be 46276Da, which corresponds to the 46278 peak. The E2 with the additional 188Da lipoic acid (46460Da) is represented by other major 46460Da peak.

Approximately 50% of the recombinant E2 polypeptides are lipoylated by the *E. coli* endogenous lipoylation pathway. In comparison to previous work done by Dr. Heath (2006), it can be concluded that inducing E2 expression with 1mM IPTG does not affect the percentage of E2 polypeptides that are lipoylated *in vivo* during recombinant expression, but it does increase the overall yield of E2 protein.

### *3.3.2.2 E2 Transacetylase coupled assay*

An assay for the E2 enzyme was developed based on the method described by Srere and Kosicki (1961). A coupling phosphotransacetylase enzyme capable of functioning above 50°C was identified from *Geobacillus stearothermophilus* (EC Number: 2.3.1.8) (Sigma-Aldrich). The temperature optimum of this enzyme was determined in order to identify an optimal temperature for the coupled assay (Figure 3.15). The enzyme has approximately 90% activity at 55°C; so for consistency with all other OADHC assays, this temperature was used.

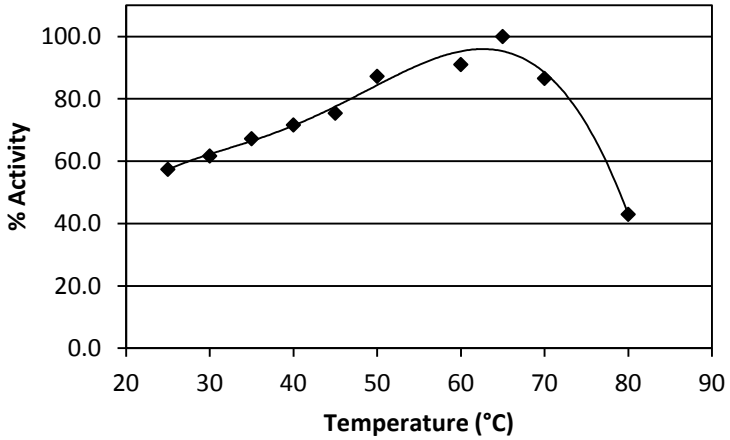



**Figure 3.15 Dependence of enzyme activity on assay temperature for the *Geobacillus stearothermophilus* phosphotransacetylase enzyme**. Phosphotransacetylase activity is expressed as a percentage of activity at 65°C

The lipoylated and non-lipoylated purified E2 enzyme samples were assayed at 55°C (Chapter 2.5.2) to determine their specific activities and $k_{cat}$ values (Table 3.2).

| E2 Enzyme | Specific activity (U/mg) | $k_{cat}$ ($min^{-1}$) |
|---|---|---|
| **Lipoylated** | 0.29 ± 0.01 | 13.6 ± 0.5 |
| **Non-lipoylated** | 1.05 ± 0.05 | 48.9 ± 2.2 |

**Table 3.2 Specific activity and $k_{cat}$ values for the recombinant E2 enzyme.** 1 Unit of enzyme activity is defined as 1µmol of acetyl-dihydrolipoamide produced per min. $k_{cat}$ was calculated using the concentrations E2 polypeptide chains.

The lipoic acid modification reduces the specific activity and hence $k_{cat}$ values of the E2 catalysed acetyl transfer reaction, this is possibly due to the presence of the endogenous lipoic acid, which may act as a competitive inhibitor of the exogenous DHlipoamide provided as an assay substrate. The E2-PTA assay measures E2 activity in the reverse

direction with respect to its function in whole complex activity, converting dihydrolipoamide and acetyl-CoA into acetyl-dihydrolipoamide + CoA (Chapter 2.5.2). Therefore, the $k_{cat}$ and specific activity values determined for this enzyme cannot be directly compared to values of the E1 and E3 enzymes; rather, this assay can be used as a method for determining relative activities such as those in the thermostability assays.

#### *3.3.2.3 E2: Thermostability*

The thermostability of the E2 protein was measured by comparing the relative amounts of soluble protein, by SDS-PAGE, after incubation across a temperature gradient for 10 min (Figures 3.16 and 3.17). The E2 protein begins to precipitate after 10 min at 85°C, with no soluble protein after 10 min at 90°C.

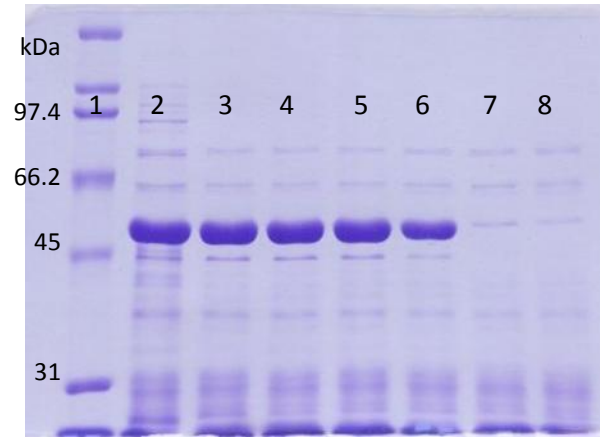



**Figure 3.16 Heat precipitation of the E2 protein - Soluble protein.** Lane 1: standard protein markers, with their Molecular weight values (kDa); lane 2: untreated sample; lanes 3 to 8: soluble protein after 10 min at 70°C, 75°C, 80°C, 85°C, 90°C, and 95°C.

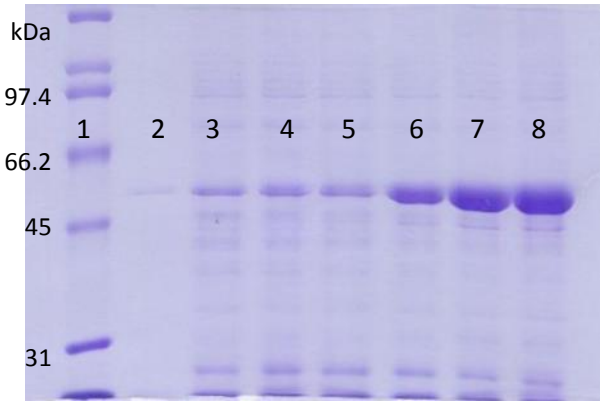



**Figure 3.17 Heat precipitation of the E2 protein - Insoluble protein.** Lane 1: standard protein markers, with their Molecular weight values (kDa); lane 2: untreated sample; Lanes 3 to 8: insoluble protein after 10 min at 70°C, 75°C, 80°C, 85°C, 90°C, and 95°C.

The thermostability of the E2 enzyme was also investigated by measuring the decrease in activity after the E2 protein was heated at various temperatures. The thermal inactivation profile of the non-lipoylated E2 enzyme is shown in Figure 3.18. After 10 min at 85°C, 20% activity had been lost, while at 90°C less than 10% activity remained after just 5 min. This loss in activity agrees with the precipitation studies, after 10 min at 85°C a small amount of precipitation occurred, whilst at 90°C the protein had completely precipitated. The E2 protein was incubated in 50mM Tris-HCl, 100mM NaCl pH 6.8 at 90°C due to precipitation of most of the E2 protein during dialysis into phosphate buffers a buffer with a lower $\Delta pK_a/\Delta T$ than Tris-HCl.

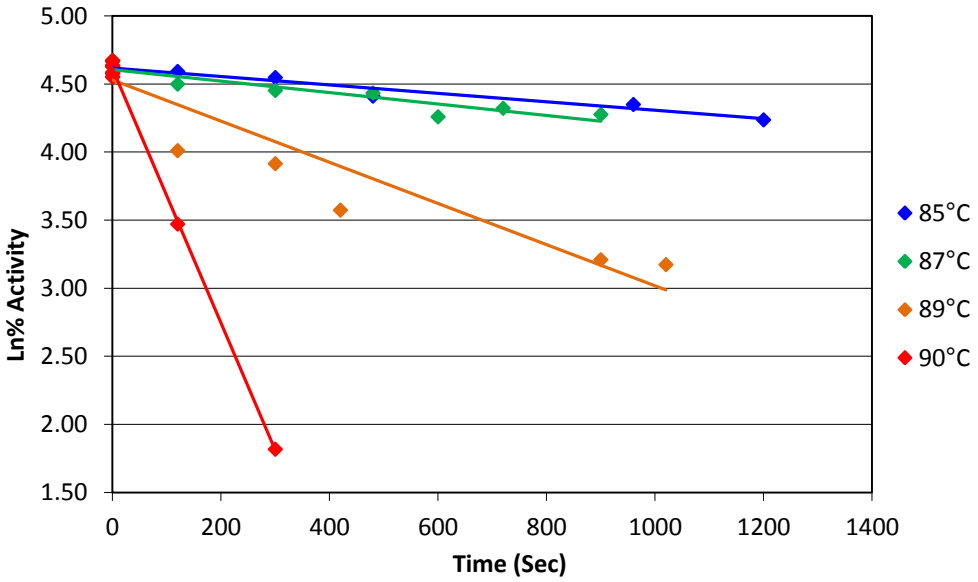



**Figure 3.18 Thermal inactivation profile of the E2 enzyme.** Each sample was heated at a defined temperature and then cooled before the remaining activity was assayed under optimal conditions. Samples were heated in 50mM Tris-HCl 100mM NaCl pH 8.8. The E2 activity before heating is taken to be 100%.

### 3.3.3 E3 Dihydrolipoamide dehydrogenase

#### *3.3.3.1 Expression and purification*

The E3 protein was purified to over 95% purity by a heat-precipitation step followed by nickel-affinity chromatography, as shown in Figure 3.19. An additional 100 kDa band is often seen at higher E3 concentrations, a proportion of E3 polypeptides appears to form inter-subunit disulphide bonds.

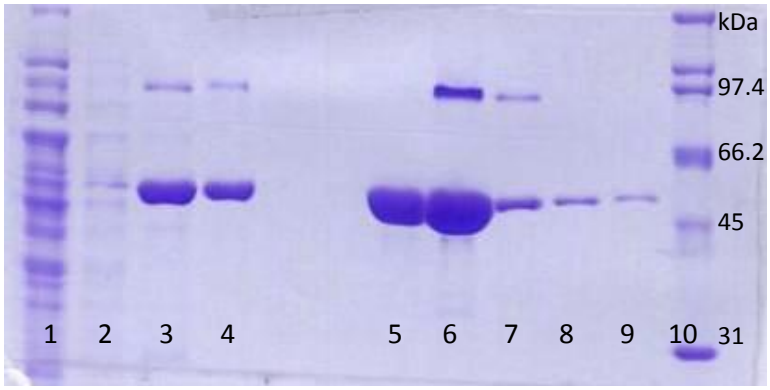



**Figure 3.191 SDS-PAGE of the purification of the E3 protein.** E3 Lane 1: flow through; lane 2: his bind buffer wash; lanes 3 and 4: 5 and 10% his elute washes; lanes 5 – 9: 1ml fractions of 100% his elute buffer; and lane 10: standard protein markers, with their molecular weight values (kDa).

#### *3.3.3.2 E3: Temperature optimum*

The temperature optimum ($T_{opt}$) of the E3 enzyme was determined by assaying the enzyme as described in Chapter 2.5.3. In this experiment, the EPPS buffer was replaced with 50mM sodium phosphate, pH 7.0, to increase consistency between conditions used in the temperature optima experiments for the other OADHC enzymes, and to reduce any temperature-dependent variation in pH (the $\Delta pK_a/\Delta T$ of phosphate is -0.0028 compared to -0.015 for EPPS). The E3 assay at pH 7.0 has a curved initial rate, so the assays were also repeated at pH 8.0 to ensure the pH was not affecting the temperature optimum of this enzyme. The temperature optima curves in Figure 3.22 show that at pH 7.0 and pH 8.0 the E3 enzyme is optimally active between 45 and 50°C.

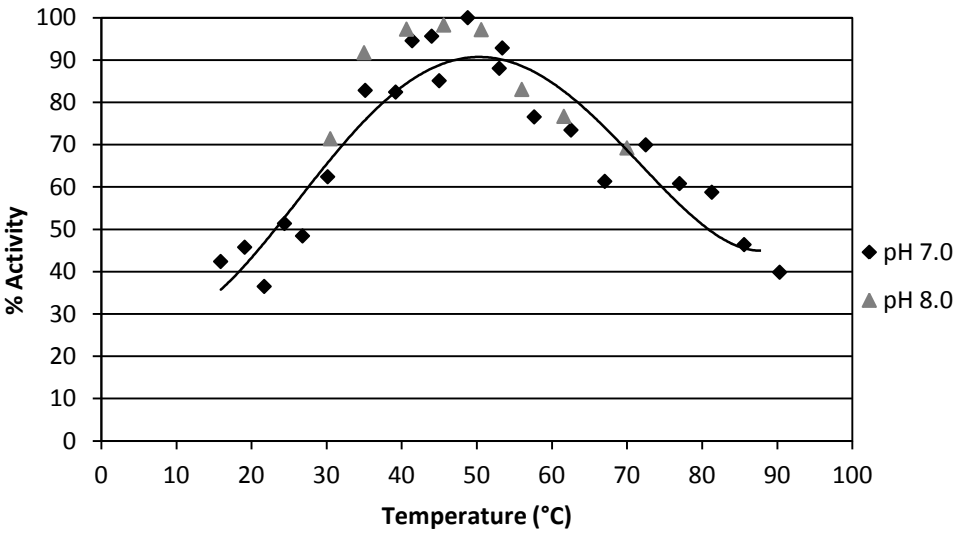



**Figure 3.20 Temperature optimum of the E3 dihydrolipoamide dehydrogenase enzyme at pH7 and pH8.** E3 activity expressed as percentage of activity of optimal recorded activity at pH7.0 and pH8.0

The E3 enzyme $T_{opt}$ curve is broader than expected, with activity only decreasing to 40% at 15°C and 90°C. The E3 enzyme has a $T_{opt}$ lower than the organism's optimal growth temperature; however, as E3 activity does not decrease rapidly after its $T_{opt}$ (65% remaining activity at 65°C) and E3 activity is not the rate limiting step of in the complex reaction (Danson et al., 1978), this lower optimal temperature may not be detrimental to overall OADHC activity.

### *3.3.3.3 E3: Thermostability*

The thermostability of the E3 protein was determined in the presence and absence of the dihydrolipoamide substrate.

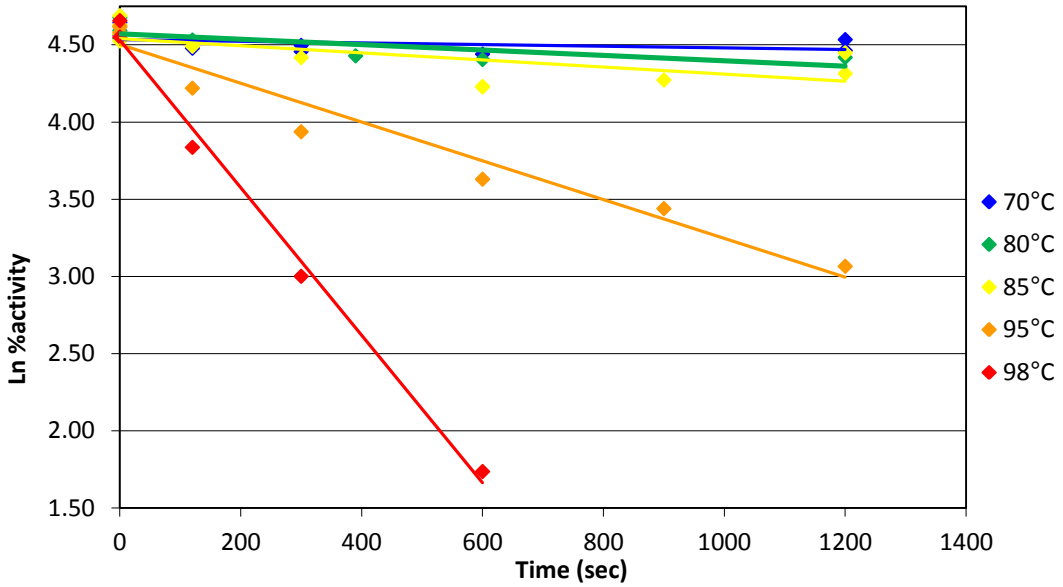



**Figure 3.21 Thermal inactivation profile of *Tp acidophilum* E3 enzyme.** Each E3 sample was heated at the specified temperature then cooled before the remaining activity was assayed. The E3 activity before heating was taken to be 100%.

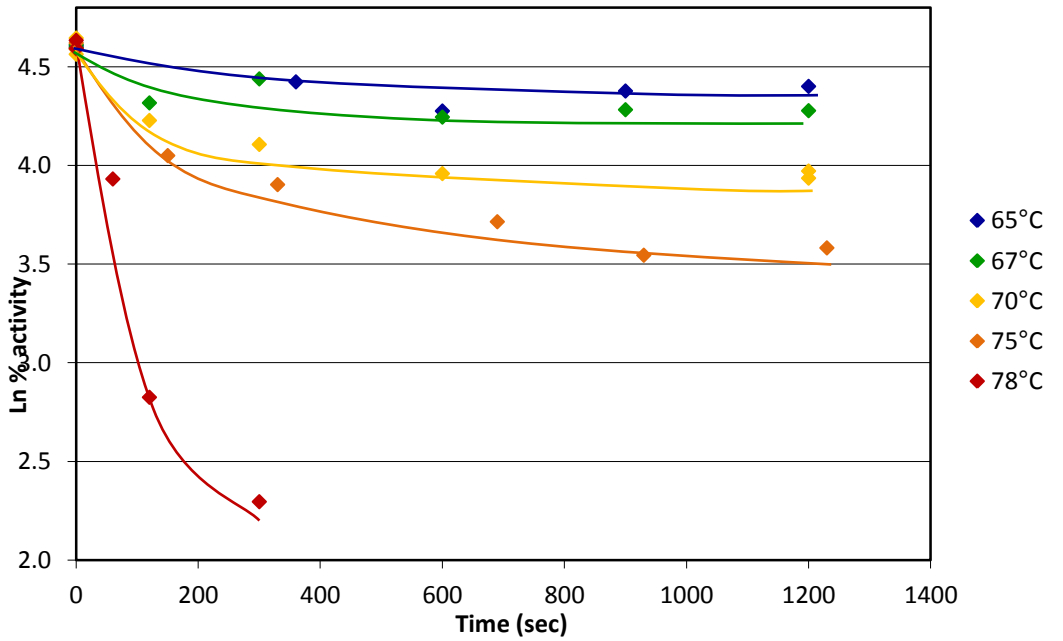



**Figure 3.22 Thermal inactivation profile of the *Tp. acidophilum* E3 enzyme with 0.4mM dihydrolipoamide.** Each E3 sample was heated as described in Figure 3.23 with the addition of 0.4mM DHlipoamide

The presence of the DHlipoamide appears to decrease the stability of the E3 enzyme. The thermal inactivation data shown in Figure 3.22 produce slightly curved lines at all temperatures which suggests that binding of the substrate influences the kinetics of the thermal inactivation of the protein. The curves could be resolved into two straight lines, there is a faster initial decrease in activity followed by a slower decrease, this suggests that more than one inactivation process is occurring.

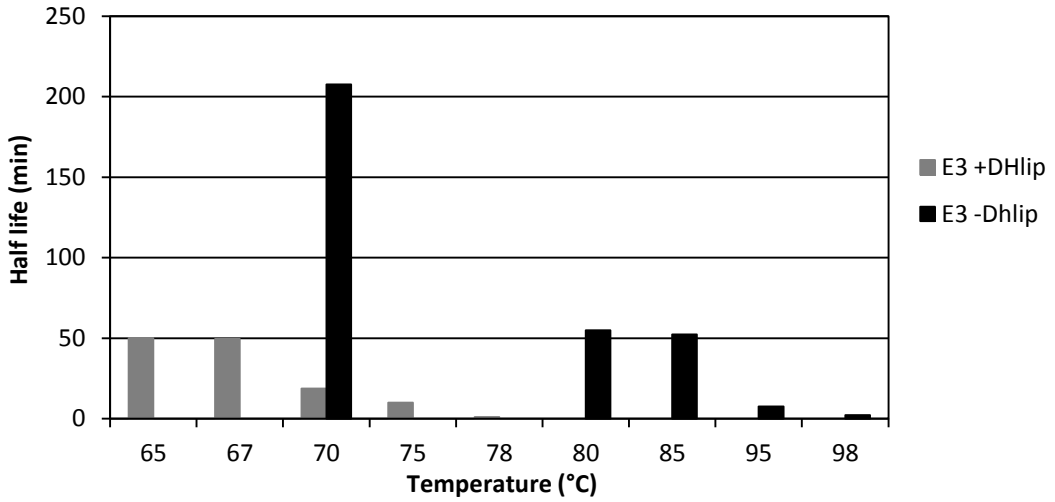



**Figure 3.23 Half life of the E3 enzyme with and without 0.4mM dihydrolipoamide at varying temperatures.**

The E3 enzyme is the most thermally stable enzyme of the all OADHC component enzymes, losing less than 10% activity after 20 min at 75°C; however the temperature optimum of the E3 enzyme is 20-25°C lower than that of the E1 enzyme. The E3 DHlip dehydrogenase is an example of an enzyme, which although stable at high temperatures, is not optimally functional. This could be due to the active site beginning to denature at a lower temperature than the rest of the protein, so that when the temperature is decreased the active site is pulled back into shape, resulting in an active protein (Daniel and Danson, 2010). After a 20 min incubation at 85°C, without substrate, the enzyme retains 80% activity; when 0.4mM DHlip is present (approximately 70% saturation) the E3 enzyme retains only 60% activity after 20 min at 65°C. Reduction of the active-site disulphide bond by the initial reaction with DHlip (Mattevi et al., 1991b) may destabilises the active site reducing catalytic activity; therefore, the thermal inactivation assays in the presence of DHlip may be closer to the condition in the temperature optima assays, as the active site disulphide will temporarily be reduced whilst catalysing the transfer of the reducing equivalents from the lipoic acid onto $NAD^+$ via a ping-pong mechanism.

### 3.3.4 Characterisation of the thermostability and temperature optimum of the branched-chain 2-oxoacid dehydrogenase multienzyme complex

#### *3.3.4.1 Tp. acidophilum OADHC assembly*

The molar concentration of each purified active enzyme assembly ($E1\alpha_2\beta_2$, E2α, and $E3\alpha_2$) was determined by measuring the absorbance at 280nm. The $E1\alpha_2\beta_2$, E2 and $E3\alpha_2$ were mixed in a molar stoichiometry of 3:1:0.1 and preheated at 55°C for 10 min, to assemble an active OADHC (Heath, 2006).

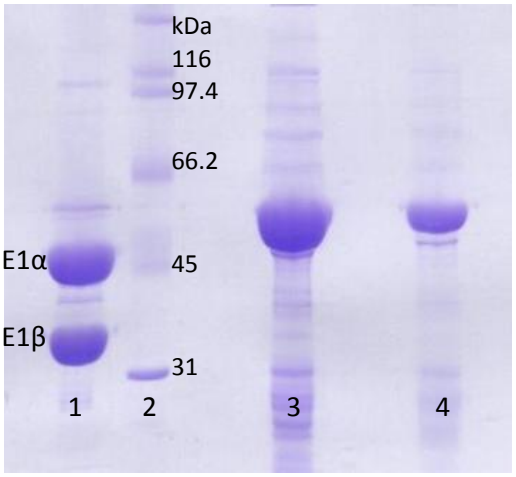



**Figure 3.24 SDS-PAGE of the purified OADHC components.** Lane 1: E1; lane 2: standard protein markers, with their molecular weight values (kDa); lane 3: lipoylated E2; and lane 4: E3.

### *3.3.4.3 Comparison of the temperature optima of the individual components and OADHC*

The complex components were mixed in a molar ratio of 3:1:0.1 (Heath, 2006) and pre-incubated to allow for cofactor binding and assembly. The temperature optimum of the OADHC was determined to be 65°C (Figure 3.25).

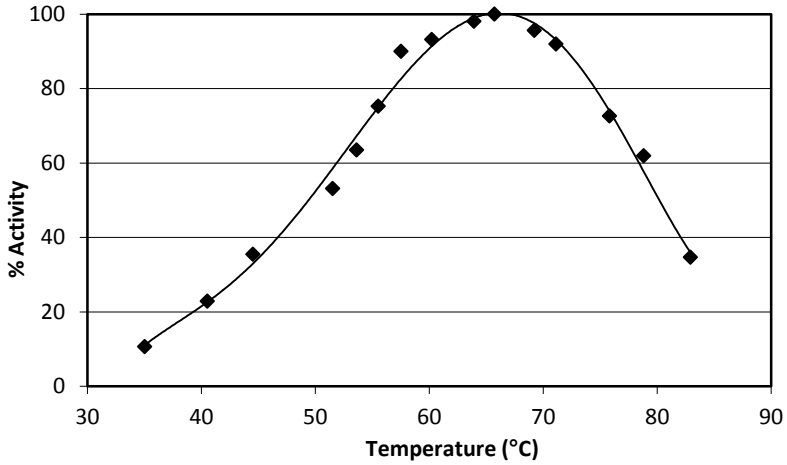



**Figure 3.25 Temperature optimum of the *Tp. acidophilum* OADHC.** The OADHC activity is expressed as percentage of activity at 65°C

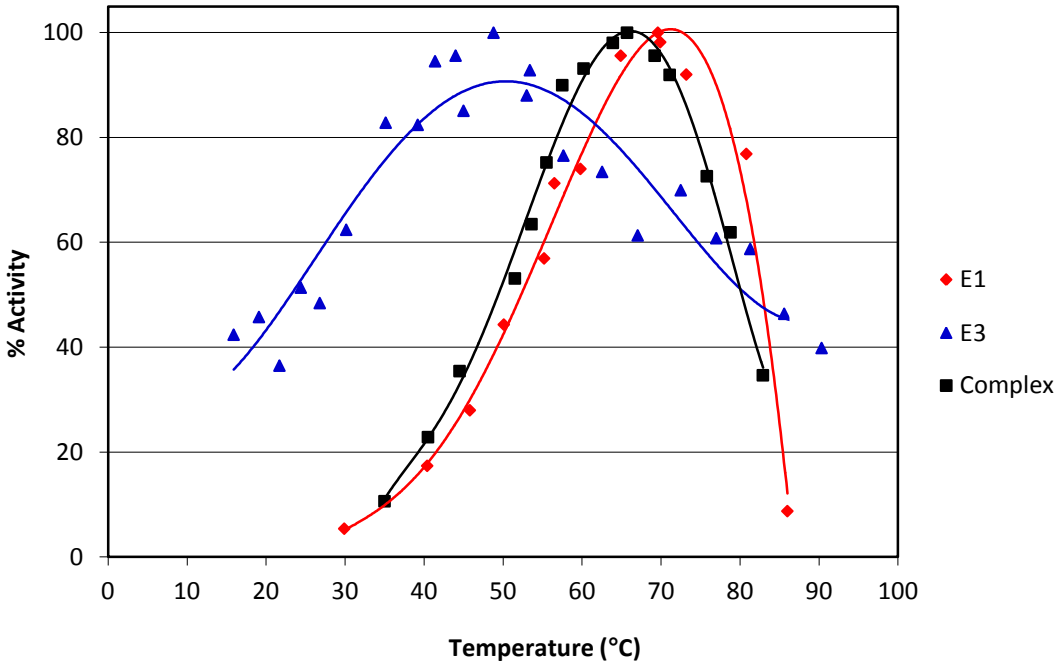



**Figure 3.26 Temperature optima profiles of the E1, E3 and the OADHC.** The E1 (red) temperature optimum is 72°C, and the E3 (blue) has a temperature optimum of 45°C. The OADHC shown in black has a temperature of 65°C. Activity of each enzyme is expressed as a percentage of the optimum recorded activity.

The temperature optimum ($T_{opt}$) of the OADHC (65°C) is similar to that of E1 (72°C), whereas, the E3 temperature optimum is significantly lower at around 45-50°C. The OADHC and E1 curves are very similar, suggesting that the $T_{opt}$ of the complex is dependent on the $T_{opt}$ of the E1 enzyme. The lower temperature optimum of the E3 enzyme does not appear to affect the $T_{opt}$ of the OADHC. The activity of the E3 enzyme reduces slowly with increasing temperature and, even though functioning at below its optimal activity, this enzyme is capable of reoxidising the lipoic acids reduced by previous reactions. To understand the effects that each individual enzyme $T_{opt}$ has on the overall complex $T_{opt}$, the temperature optimum of the E2 enzyme should ideally be known; however, this could not be determined due to the coupled nature of theE2-PTA assay.

### *3.3.4.2 Comparison of the thermal inactivation of the individual enzymes and the assembled OADHC*

The OADHC was assembled in a stoichiometry of 3:1:0.5 ($E1\alpha_2\beta_2$:E2:$E3\alpha_2$). The complex components were mixed and diluted 2-fold in 200mM sodium phosphate, pH 7.0, 10% glycerol, 2mM $MgCl_2$, and 0.2mM TPP, before a 10-min incubation at 55°C to permit the assembly to reach completion.

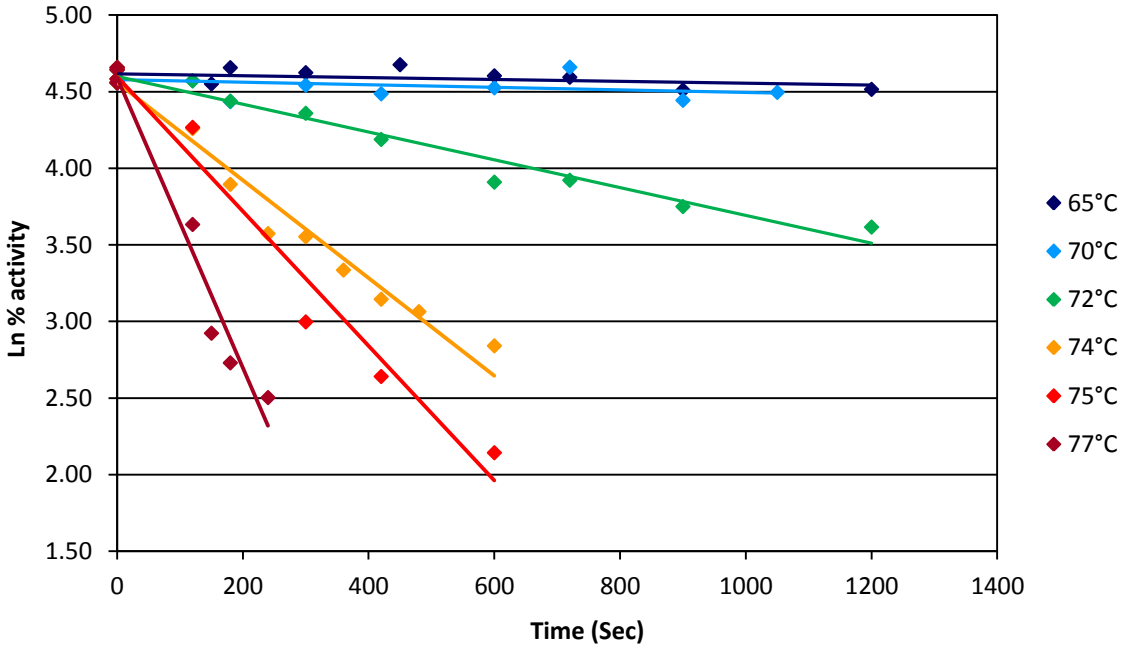



**Figure 3.27 Thermal inactivation profile of the OADHC**. The OADHC activity before heating is taken to be 100%.

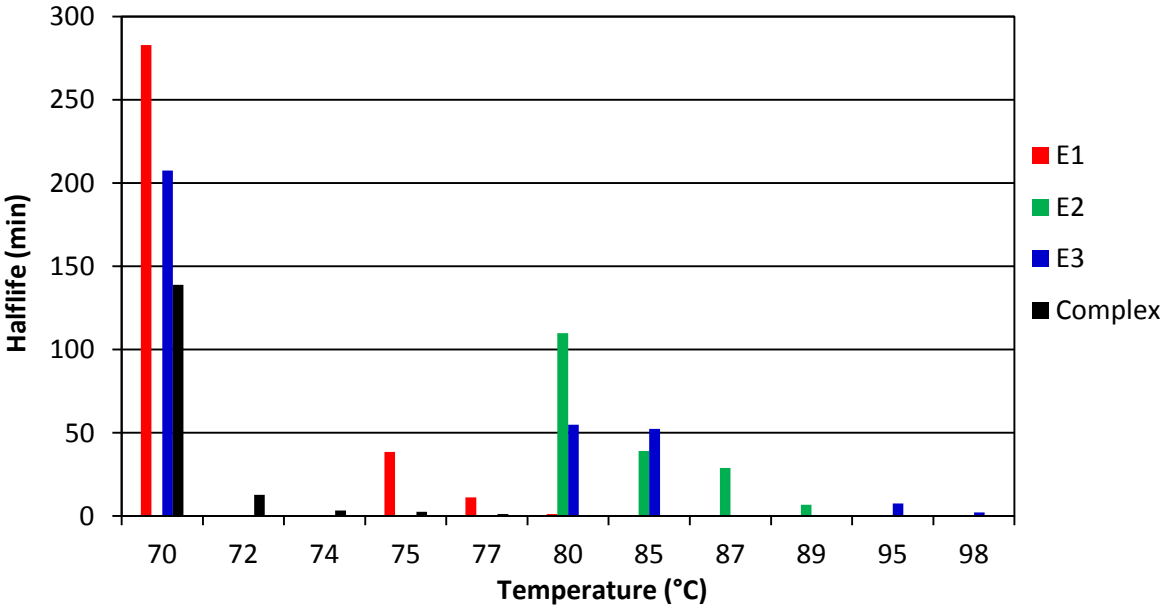



**Figure 3.28 Bar chart illustrating the half lives of the OADHC and its component enzymes**

The individual enzymes appear to be more stable than the assembled OADHC, suggesting that factors other than the thermostability of the individual enzymes are affecting the stability of the complex; this may point to the non-covalent interactions between the complex components being the bonds that are first disrupted by increased temperatures.

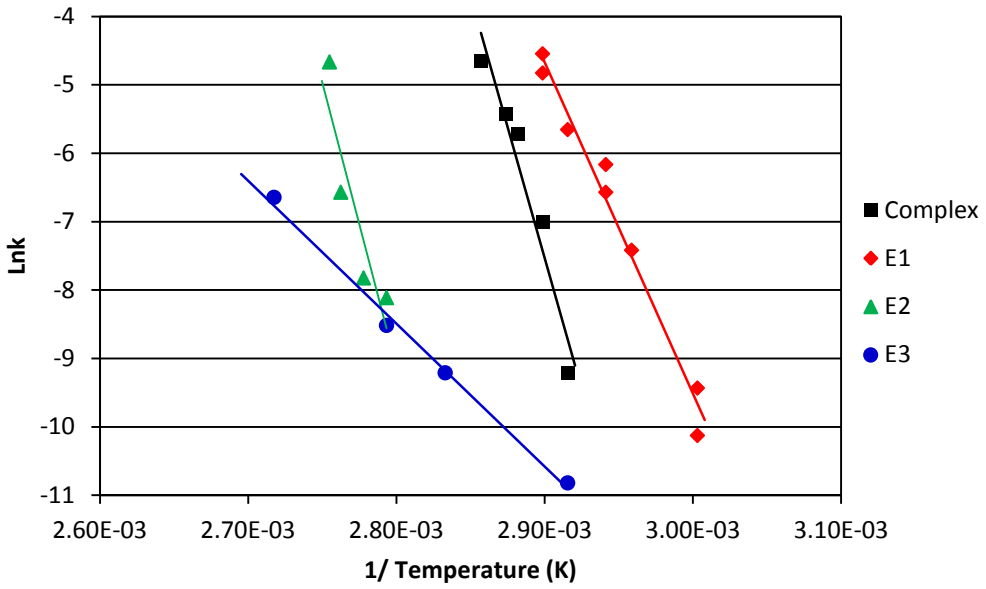



**Figure 3.29 Arrhenius plot of the thermal inactivation data of the individual enzymes and assembled OADHC.**

The Arrhenius equation can be rearranged into a straight line equation as shown below:

Arrhenius Equation: $k = A\, e^{-(Ea/RT)}$

and therefore: $Ln(k) = -\frac{Ea}{R} \cdot \frac{1}{T} + Ln(A)$

Where k = the rate of inactivation, Ea = activation energy of inactivation, R= universal gas constant, T = temperature (K), and A = Arrhenius constant.

The activation energy of inactivation (Ea) and the Arrhenius constant (A) are compared below in Table 3.3 for the individual proteins and the OADHC.

| Enzyme | -Ea/R | Ea (KJ/mol) | LnA | A (Units) |
|---|---|---|---|---|
| **E1** | 48400 | 402 | 136 | $1.2 \times 10^{59}$ |
| **E2** | 83200 | 692 | 224 | $1.9 \times 10^{97}$ |
| **E3** | 20900 | 174 | 50 | $5.2 \times 10^{21}$ |
| **Complex** | 76800 | 639 | 215 | $2.4 \times 10^{93}$ |

**Table 3.3 Activation energy of inactivation and Arrhenius constant of the individual enzymes and the OADHC.** Values were calculated from the data shown in Figure 3.29.

The Ea values calculated appear to suggest that the E3 enzyme is the least stable of the OADHC proteins assayed with the E2 and OADHC being the most stable; this contradicts the half-life data shown in Figure 3.27 in which E3 is the most stable and the OADHC the least.

Transition state theory describes the rate of thermal inactivation energy by the following equation:

$$k = \frac{k_B T}{h} e^{-\left(\Delta G^{\circ *}/RT\right)}$$

and therefore: $$\mathrm{Ln}(k) = \frac{-\Delta H^{o*}}{R} \times \frac{1}{T} + \left[\mathrm{Ln}\left(\frac{K_B T}{h}\right) + \frac{\Delta S^{o*}}{R}\right]$$

Where k = the rate of inactivation, R= universal gas constant, T = temperature (K), $K_B$ = Boltzmann's constant, G°* = Standard Gibbs free energy of inactivation, H = enthalpy of inactivation, and S = entropy of inactivation.

In this transition state theory equation the Arrhenius activation energy (Ea) now becomes a measure of the enthalpy of inactivation (ΔH), and the Arrhenius constant, a measure of the entropic contributions to inactivation (ΔS). The rate of thermal inactivation of a protein is dependent on entropic and enthalpic contributions; the rate of inactivation can be reduced by either increasing ΔH or decreasing ΔS. The entropic and enthalpic contributions determining the rate of inactivation of the E3 are different from the other enzyme with the rate influenced by smaller values of both the ΔS and ΔH. The entropic and enthalpic contributions to the thermostability of the individual enzymes and the OADHC could be further investigated by isothermal titration calorimetry.

### 3.3.5 Stoichiometry of the multienzyme components

#### *3.3.5.1 Complex stoichiometry assays*

The affect on OADHC activity of different E1:E2:E3 stoichiometries were determined by mixing differing ratios of E1 and E3 to E2 and comparing the rate of whole complex activity (Figure 3.31).

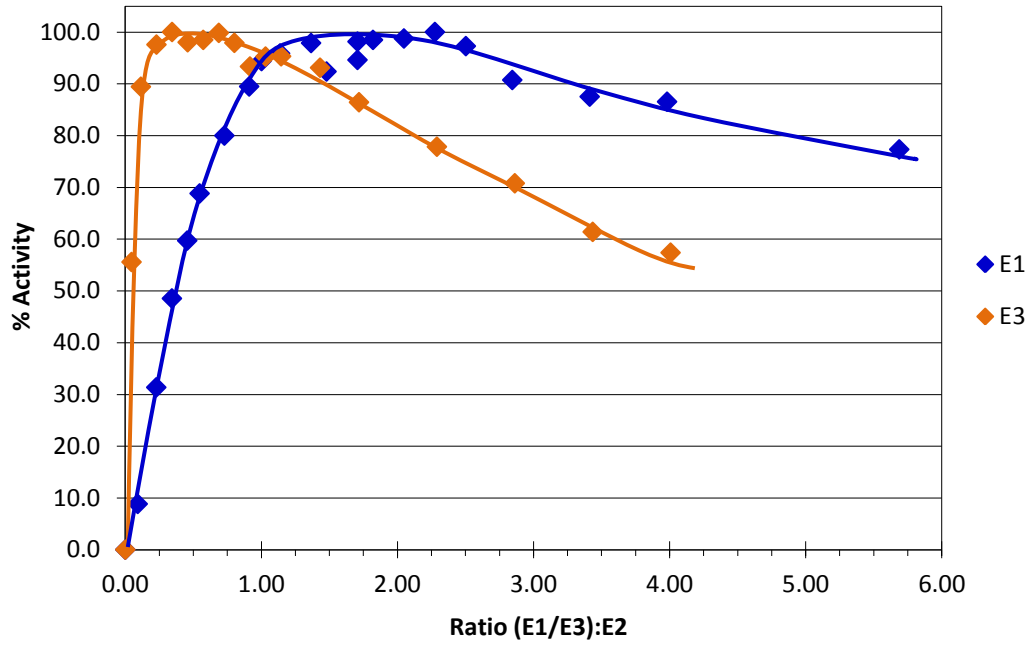



**Figure 3.30 Graph showing the percentage activity of various stoichiometries of E1 and E3 to E2.** The ratio of E1$\alpha_2\beta_2$ and E3$\alpha_2$ were varied independently with respect to the concentration of E2$\alpha$ with either the E1:E2 or the E3:E2 ratio set at 1:1. The maximum activity of E1 and E3 is taken as 100% activity.

Optimal OADHC activity was achieved at a stoichiometry of 1.75:1:0.5 E1$\alpha_2\beta_2$:E2:E3$\alpha_2$. This is different from the value of 3:1:0.1 determined by Dr. C Heath (2006); the removal of the his-tags from the E1$\beta$ may have affected interactions between the E1 and the peripheral subunit binding domain. Optimal OADHC activity was achieved with three times more E1 compared to E3. The E1 enzyme has previously been shown to be the rate limiting step in the *E. coli* PDHC by determining the rate of E2 acetylation in the absence of CoA and comparing it to the rate of whole complex activity (Danson et al., 1978). The ratio of 3:1 E1:E3 allows maximal activity by balancing an increase in the amount of rate limiting E1, with a smaller amount of E3, an enzyme that is capable of reoxidising any available lipoyl domains as quickly as they are reduced. A second explanation for

achieving optimal activity with three-fold more E1 than E3 is that the E2-PSBD may have a higher affinity for E3 than E1, in which case more E1 might be needed to out compete the E3 for PSDB binding sites. In order to determine which explanation is correct OADHC activity would need to be compared with the ratio of E1 and E3 actually bound to the E2-PSBD.

### *3.3.5.2 Complex gel filtration*

The stability of the assembled complex was investigated by gel filtration. The OADHC components were mixed in a molar stoichiometry of 1:1:1 ($E1\alpha_2\beta_2$:E2:$E3\alpha_2$) and diluted 1:1 in 50mM potassium phosphate, pH 7.7, 1mM $MgCl_2$, and 2mM 3-methyl-2-oxopentanoic acid, before being incubated at 55°C for 10 min to allow assembly and cofactor binding. The column was also equilibrated with the same buffer.

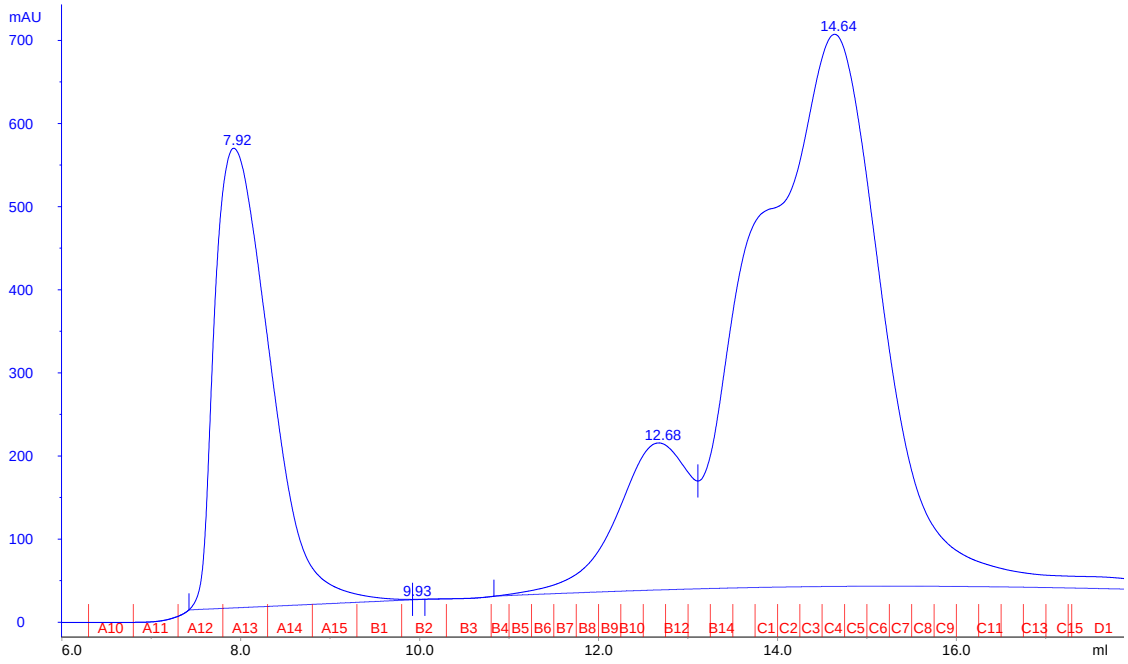



**Figure 3.31 Gel filtration trace of the OADHC.** The $A_{280}$ readings are shown in blue, and the fraction numbers shown along the x-axis in red. The column was equilibrated with 50mM potassium phosphate, 1mM $MgCl_2$, 2mM 3-methyl-2-oxopentanoic acid and pH7.7

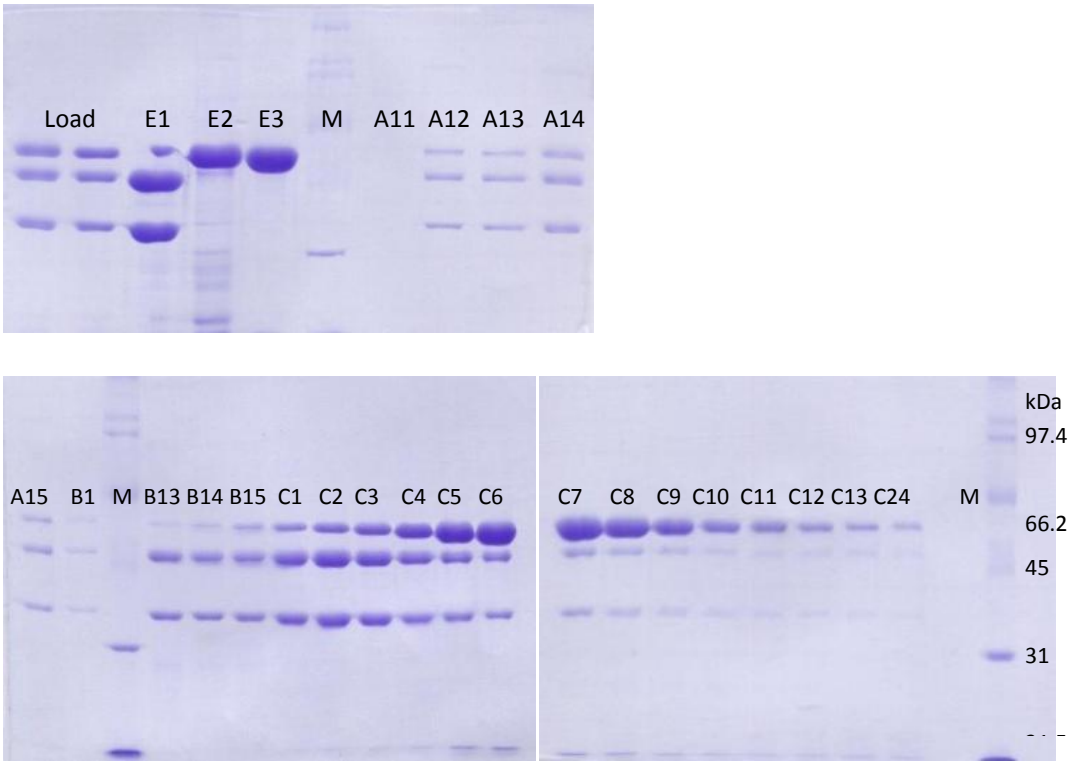



**Figure 3.32 SDS-polyacrylamide gels of the Superdex 200 gel filtration fractions.** Each lane is labelled with the corresponding fraction number. The standard protein markers, with their molecular weight values (kDa), are shown in the lanes labelled M. Lanes E1, E2 and E3 are samples of the purified recombinant proteins used in the assembly process.

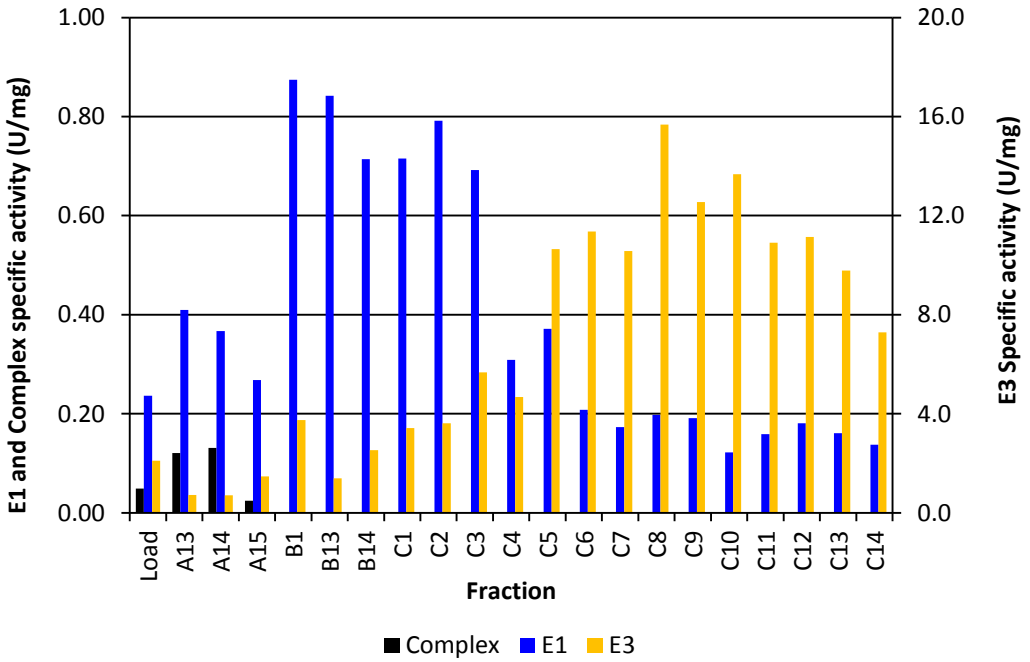



**Figure 3.33 A bar chart illustrating the specific activity of the E1, E3 and whole complex in each fraction after gel filtration.** The specific activity is shown in U/mg where 1Unit is equivalent to 1µmol of product per min, and the activity is shown per mg of total protein in each sample.

The 1:1:1 ($E1\alpha_2\beta_2$:E2α:$E3\alpha_2$) stoichiometry should saturate the E2, as there is twice the number of E1 and E3 enzymes compared to the E2, each E2 polypeptide binds can one E1 or E3 in a mutually exclusive manner. An excess of unbound E1 and E3 would be expected to elute from the gel filtration column as the E2 polypeptides would be able to bind only half the E1 or E3 enzymes; however, 92.4% of the E1 and 98.9% of the E3 activity is found in the unbound fractions, with only 7.6% and 1.1% E1 and E3 activities present in the complex fractions. The OADHC formed by mixing the components in a stoichiometry of 1:1:1 has a higher specific activity than any lower ratio, but it would appear, that this mixture contains a large amount of E1 and E3 enzymes that either do not bind to the E2 or only bind weakly.

Key residues involved in the interaction between *Geobacillus stearothermophilus* E3 and the E2-PSBD have been well characterised by Mande et al. (1996); a sequence alignment of the *G. stearothermophilus* and *Tp. acidophilum:* E2-PSBD and E3, identifies any differences between these key amino acids. Table 3.3 summarises the results of the alignment; the majority of the key residues are conserved or similar, therefore this alignment does not indicate why the *Tp. acidophilum* E2-PSBD and E3 binding appears to be weak. This alignment only compares sequence similarity, so does not take into account any structural changes which may affect the positions of the side chains involved in these interactions.

| | *G. stearothermophilus* | | | | *Tp. acidophilum* | |
|---|---|---|---|---|---|---|
| | **E2-PSBD** | | **E3** | | **E2-PSBD** | **E3** |
| Polar | Ser 133 | $O_{\gamma}$ | Glu A437 | $O_{\varepsilon 1}$ | ✗ - Ala | ✗ -Met |
| | Ser 133 | N | Asp A438 | $O_{\delta 1}$ | ✗ - Ala | ✔ |
| | Arg 135 | $N_{\varepsilon}$ | Asp B344 | $O_{\delta 2}$ | ✔ | ✗ - *Glu* |
| | Arg 135 | $N_{\eta}1$ | Glu B431 | O | ✔ | ✔ |
| | Arg 135 | $N_{\eta}2$ | Glu B431 | $O_{\varepsilon 2}$ | ✔ | ✔ |
| | Lys 136 | $N_{\zeta}$ | Glu A437 | $O_{\varepsilon 2}$ | ✗ - Arg | ✔ |
| | Arg 139 | $N_{\eta}2$ | Asp B344 | $O_{\delta 1}$ | ✔ | ✗ - *Glu* |
| | Arg 139 | $N_{\eta}2$ | Ala B432 | O | ✔ | ✔ |
| Non-Polar | Met 131 | $C_{\beta}$ | Asp A438 | $C_{\gamma}$ | ✗ - Ser | ✔ |
| | Pro 132 | $C_{\delta}$ | Leu A441 | $C_{\delta 1}$ | ✔ | ✔ |
| | Pro 132 | $C_{\beta}$ | Leu A441 | $C_{\beta}$ | ✔ | ✔ |
| | Pro 132 | $C_{\gamma}$ | Glu B431 | $C_{\gamma}$ | ✔ | ✔ |
| | Ser122 | $C_{\gamma}$ | Asp A438 | $C_{\gamma}$ | ✗ - Ala | ✔ |
| | Arg 139 | $C_{\beta}$ | Ala A342 | $C_{\beta}$ | ✔ | ✔ |
| | Arg 139 | $C_{\beta}$ | Ala B342 | $C_{\beta}$ | ✔ | ✔ |
| | Glu 140 | $C_{\gamma}$ | Pro B334 | $C_{\gamma}$ | ✔ | ✗ - Asp |

**Table 3.3 Comparison of the residues involved in polar and non-polar interactions between the *G. stearothermophilus* E3 and peripheral subunits binding domain (Mande et al., 1996) compared to *Tp. acidophilum* E3 sequence.** A sequence alignment has been used to determine if the amino acids involved in the interaction between the *G. stearothermophilus* E3 and PSBD are found in *Tp. acidophilum*. The presence of a conserved residue is indicated by a tick; crosses have been used to indicated when the residue is not identical and in each case amino acid in that position has been stated.

## 3.4 Conclusion

The results in this chapter demonstrate that the recombinant E1, E2 and E3 enzymes form a complex that catalyses the conversion of branched chain 2-oxoacids to their corresponding acyl-CoAs at temperatures within the growth range of *Tp. acidophilum*. This result, along with the identification of the *Tp. acidophilum* E2 lipoylation mechanism (LplA-CTD) (Posner et al., 2009), indicates that the native E1, E2 and E3 enzymes could be functional, if their genes were transcribed and translated by the organism. E3 enzyme activity has been detected in *Tp. acidophilum* cell lysates but no whole complex activity has been detected in this, or any other archaeon (Danson et al., 1984). The lack of whole complex activity could be explained by the apparent weak interaction between the E1 and E3 enzymes with the E2 core. A proteomics studies by Sun et al. (2007) that identified large complex proteins from *Tp. acidophilum* cell lysates found that the E2 and E1 proteins formed a large protein complex no E3 was present. The results described in this chapter along with the study by Sun et al. (2007) suggest that the OADHC genes are transcribed and translated in T*p. acidophilum* resulting in functional individual enzymes, but this enzymes do not strongly associate in cell lysates so no active OADHC is formed.

It has also been shown that there is no change in the expression of the *Tp. acidophilum* OADHC genes under aerobic and anaerobic conditions, but under anaerobic conditions there is up-regulation of the ferredoxin oxidoreductase genes (Sun et al., 2011). The OADHC in *Tp. acidophilum* may be regulated by other conditions that have not yet been determined, such as changes to growth substrate, or in response to different respiratory terminal electron acceptors. The presence of nitrate, a respiratory electron acceptor, resulted in a 16-fold up-regulation of the OADHC genes in *Pyrobaculum aerophilum* compared to anoxic conditions, although the up-regulation was significantly less compared to aerobic conditions (Cozen et al., 2009). It would be interesting to compare the expression levels of the central metabolism genes after growth of *Tp. acidophilum* on different types and levels of growth substrates.

# Chapter 4: Cloning, expression and characterisation of the E2 catalytic domain

## 4.1 Introduction

The E2 dihydrolipoamide transacetylase enzyme is composed of three functional domains separated by flexible linker regions (Figure 4.1).

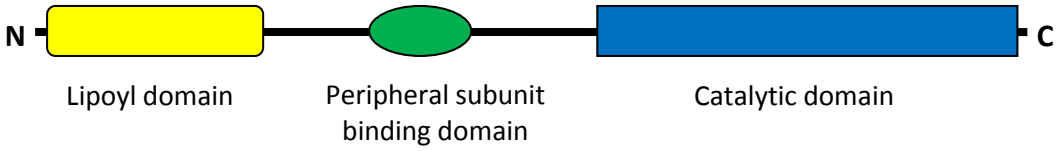



**Figure 4.1 The *Tp. acidophilum* E2 protein domain structure.** The N-terminal 8.5 kDa lipoyl domain is shown in yellow, the 6 kDa peripheral subunit (E1 and E3) binding domain is shown in green, and the C-terminal 25 kDa catalytic domain in blue.

The thermostability of the catalytic domain is crucial to the stability of the 2-oxoacid dehydrogenase multienzyme complex (OADHC). A high-resolution structure of the fully-assembled E2 protein may provide insights into the thermostability of this protein complex. Ideally, an X-ray crystallographic structure would provide these details; however, due to the flexible nature of this multi-domain protein, the whole E2 protein is unlikely to crystallise. The experiments detailed in this chapter, aimed to clone and express the E2 catalytic domain (E2cat) by removing the flexible regions, the peripheral subunit binding domain (PSBD) and the lipoyl domain. The E2cat protein was then compared to the full-length E2, to determine the affects this truncation had on enzymatic activity and thermostability.

## 4.2 Methods

### 4.2.1 Cloning of the E2 catalytic domain gene

The location of the forward primer was chosen based on sequence alignments with other E2 sequences, for which crystal structures were available. Figure 4.2 shows the known sequences of three E2 catalytic domain structures along with the whole *Tp. acidophilum* E2 sequence.

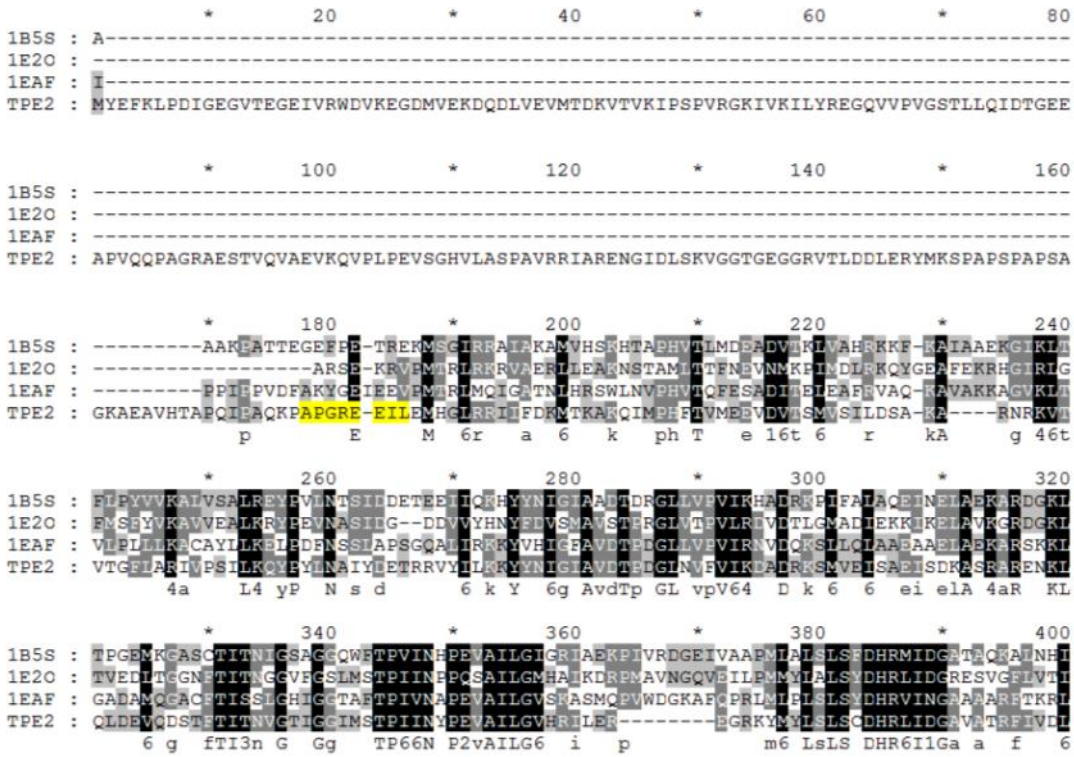



**Figure 4.2 A protein sequence alignment of the E2 catalytic domains with known structures and the *Tp. acidophilum* E2.** The sequences are labelled by their PDB access code: 1B5S, *Geobacillus stearothermophilus* PDHC*; 1E2O, Escherichia coli* 2-oxoglutarate dehydrogenase complex; and 1EAF, *Azotobacter vinelandii* PDHC. TPE2 is whole *Tp. acidophilum* E2 protein sequence. The region highlighted in yellow is the *Tp. acidophilum* sequence complementary to forward primer.

The E2cat primers are shown below; the sequence complementary to the E2 gene is in **bold**, the restriction sites (*Xho*I and *Nde*I) are in *italics* and any additional bases added in grey.

E2 Cat Forward $^{5'}$*CAT ATG* **GCA CCT GGG AGA GAA GAG ATC CTG G**$^{3'}$

E2 Reverse $^{5'}$CCG *CTC GAG* **TCA GAT CTC GTA GAT TAT AGC GTT CGG**$^{3'}$

The E2cat gene was PCR amplified and ligated via the cloning vector pGem-T Easy (Promega, U.K.), into the expression vectors pET24a and pET28a (Novagen, Merck, U.K.) as described in Chapter 2.2.

### 4.2.2 Expression and purification of E2cat protein

#### *4.2.2.1 E2cat protein expression*

The E2cat protein was expressed without a his-tag, from pET24a, following the method described in Chapter 3.2.1. The cell lysate was incubated at 55°C for 10 min and centrifuged to remove the heat-precipitated proteins.

#### *4.2.2.2 Polyethylenimine DNA precipitation*

The *E. coli* genomic DNA was precipitated before anion exchange chromatography, by the titration of the cell lysate with polyethylenimine (PEI) (BDH Ltd.). A range of concentrations (0.1-0.6% (v/v)) were tried to determine the amount required to precipitate the DNA whilst not affecting the solubility of the E2cat protein. The cell lysate containing PEI was incubated on ice for 1 h, and then centrifuged at 16000 x g for 15 min, to remove all precipitated DNA. The supernatant was dialysed overnight into 50mM Tris-HCl, pH8.8.

#### *4.2.2.3 Anion exchange chromatography*

Anion exchange chromatography was carried out using the Äkta FPLC system (Amersham Biosciences, U.K.), using two 5ml Q-Sepharose Hi-Trap column equilibrated in 50mM Tris-HCl, pH 8.8. The E2cat protein was eluted from the column using a 0-1M NaCl gradient in the same buffer.

#### *4.2.2.4 Gel filtration purification and size determination*

The final purification step was carried out as described in Chapter 2.3.3.3.

### 4.3.3 Characterisation of E2cat

#### *4.3.3.1 Mass spectrometric analysis*

The mass spectrometry experiments in this chapter were performed and analysed by Trevor M. Gibson, at The BioCentre Facility, University of Reading. A sample of purified E2cat protein at 2mg/ml was dialysed against Milli-Q water overnight at 4°C. Acetic acid was added to a final concentration of 0.1% (v/v) before samples were loaded onto the MicrOTOF$^{TM}$ mass spectrometer coupled with an electrospray source (ESI-TOF) (Bruker Daltonics Bremen, Germany.).

#### *4.3.3.2 E2 dihydrolipoamide transacetylase assays and thermal inactivation*

The E2 assay and thermal inactivation experiments are described in Chapters 2.5.2 and 3.2.6, respectively. The purified E2cat protein in 50mM Tris-HCl, pH 8.8, and 100mM NaCl, was diluted in 50mM Na phosphate, pH 7.0, in a ratio of 1µl:4µl (protein:buffer) before each thermal inactivation incubation.

#### *4.3.3.3 Dynamic light scattering*

Dynamic light scattering (DLS) was used to determine the hydrodynamic diameter of the E2cat protein; all measurements were performed using a Zetasiser Nano S from Malvern Instruments Ltd. (Malvern, U.K.). 50µl of purified sample was loaded into a low volume quartz cuvette and heated to 25°C or 55°C. The thermostability was studied by performing measurements at various temperatures; the protein was heated to the required temperature and held for 2-5 min before each measurement was taken. The relative molecular mass of the protein was predicted based on the known hydrodynamic diameter of other globular proteins, using the Protein utilities tools within the Zetasiser Nano S program.

## 4.3 Results and discussion

### 4.3.1 Cloning and expression

The E2cat gene was PCR amplified from *Tp. acidophilum* genomic DNA; the gene was cloned using *Nde* I and *Xho* I into pET28a and pET24a, resulting in recombinant proteins with and without an N-terminal his-tag. The presence of the correct gene was confirmed by a restriction digest of the pET vectors (Figure 4.3) and DNA sequencing by Source Bioscience (U.K.).

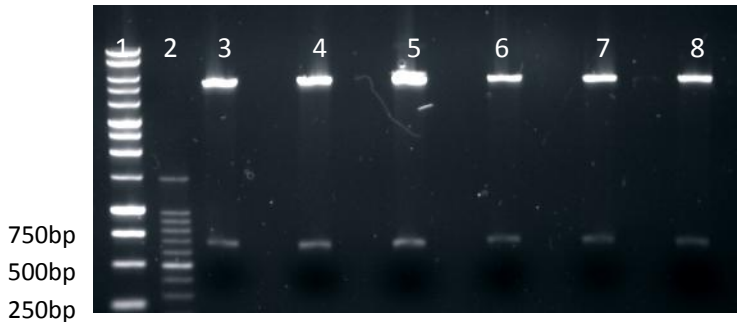



**Figure 4.3 Restriction digest of pET28a (lanes 3-7) and pET24a (lane 8)** with *Nde* I and *Xho* I to release the 678bp E2cat gene. The markers shown in lanes 1 and 2 are 1Kb ladders and 100bp ladder (Promega, U.K.).

The E2cat protein was completely insoluble with an N-terminal his-tag; whereas, the non-tagged protein was partially soluble. The pET24a E2cat protein without his-tag was used for all further experiments.

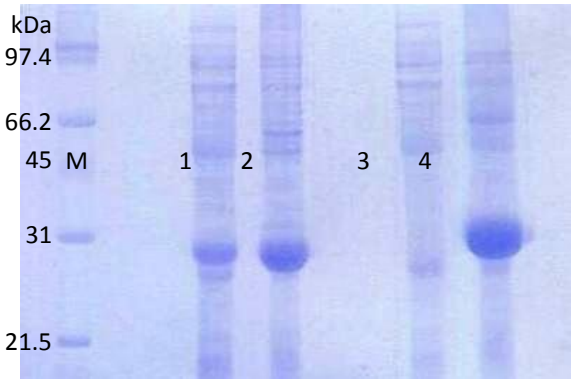



**Figure 4.4 SDS-PAGE of E2cat expression trial.** Lanes 1 and 2 are the soluble and insoluble un-tagged E2cat protein expressed from pET24a. Lanes 3 and 4 show the soluble and insoluble his-tagged E2cat protein expressed from pET28a. The standard protein markers, with their molecular weight values (kDa), are shown in the lanes labelled M.

### 4.3.2 Purification of the E2 catalytic domain

#### *4.3.2.1 Polyethyleneimine DNA precipitation*

Initial purification attempts using anion exchange chromatography and gel filtration resulted in E2cat protein which was contaminated with DNA/RNA; therefore a polyethylenimine (PEI) DNA/RNA precipitation step was added between the heat precipitation and anion exchange steps. In each purification, the cell lysate was titrated with PEI to determine the optimal volume required to precipitate the DNA without precipitating the E2cat protein (Figures 4.5 and 4.6), usually between 0.2-0.3% (v/v) PEI was used.

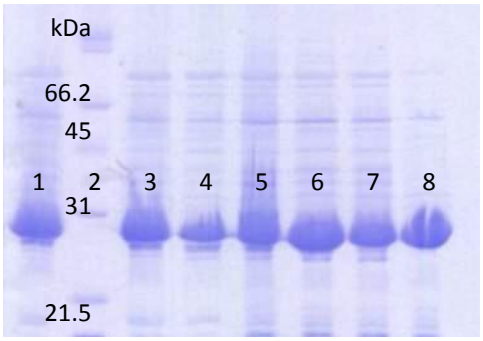



**Figure 4.5 SDS-PAGE of the PEI DNA precipitation step during the purification of the recombinant E2cat protein.**
Lane 1: the cell lysate after heat treatment; lane 2: standard protein markers, with their molecular weights (kDa); lanes 3 -8: cell lysate incubated

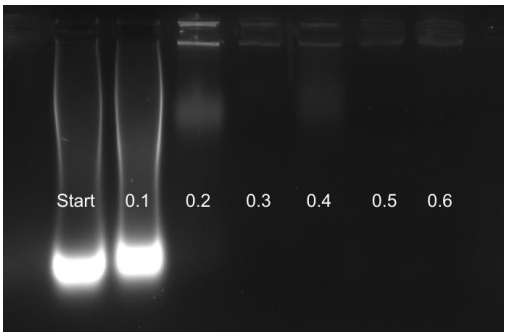



**Figure 4.6 Agarose DNA gel of the PEI DNA precipitation step during the purification of the recombinant E2cat protein.** The lanes are labelled with the percentage (v/v) of PEI added to the cell lysate.

### *4.3.2.2 Anion exchange Chromatography*

Following the PEI precipitation, the E2cat protein was further purifed by anion exchange chromatography (Figures 4.7 and 4.8).

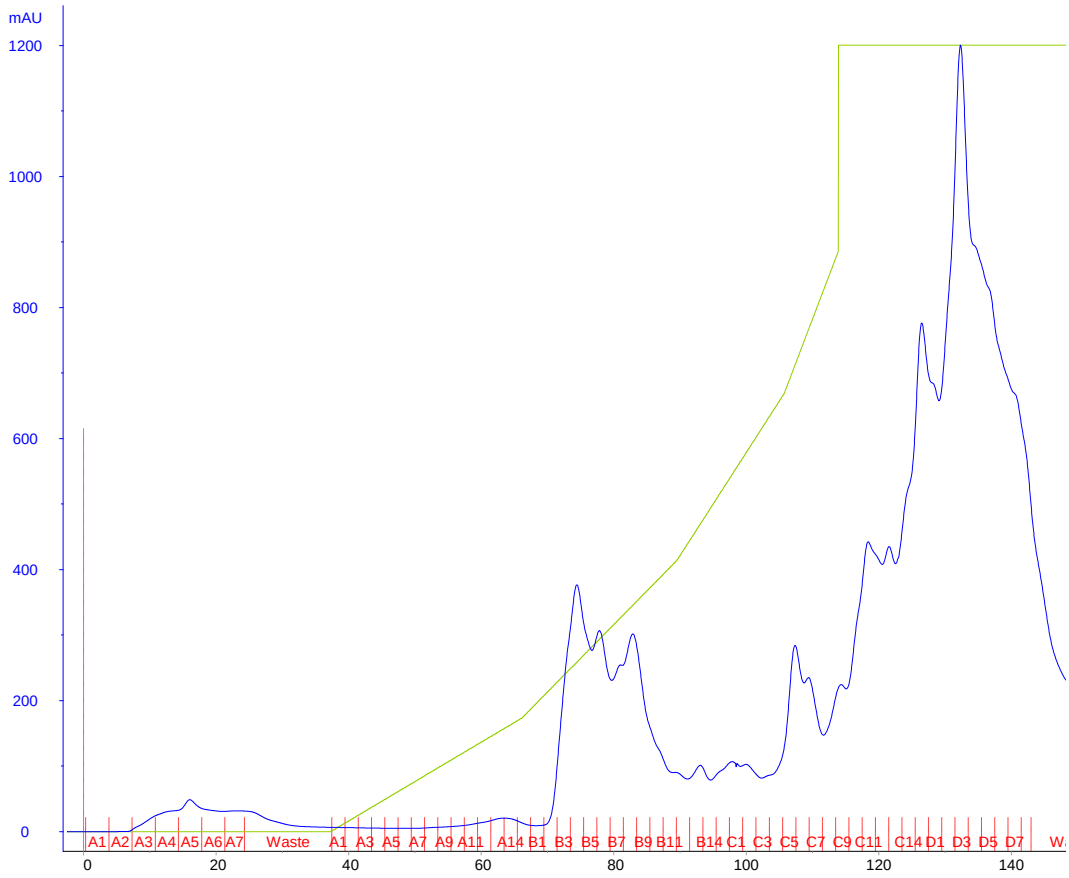



**Figure 4.7 An E2cat purification anion exchange trace.** The $A_{280}$ readings are shown in blue; the concentration gradient of NaCl from 0-1M is shown in green; and the 2ml fraction numbers are shown along the x-axis.

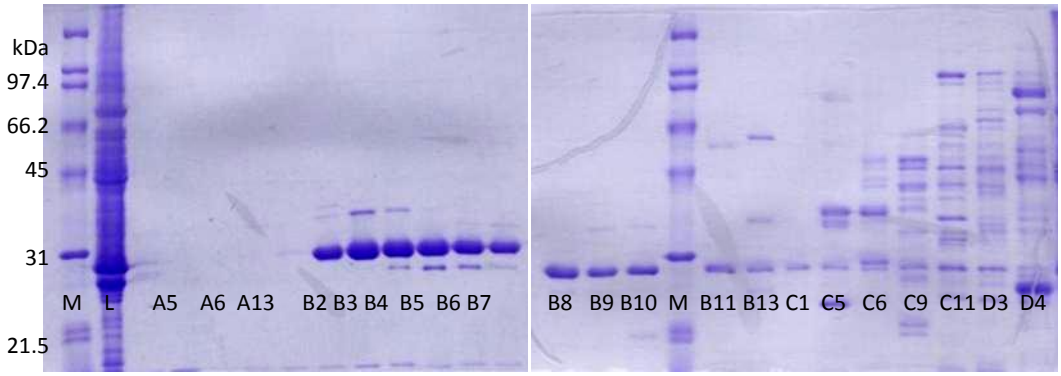
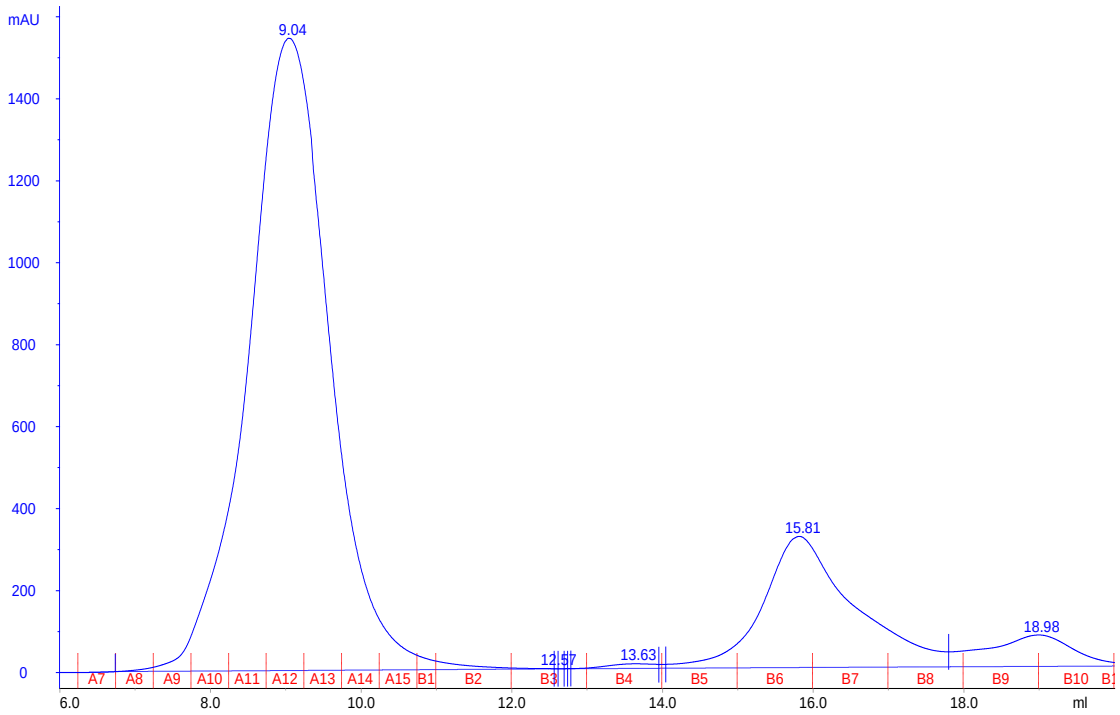


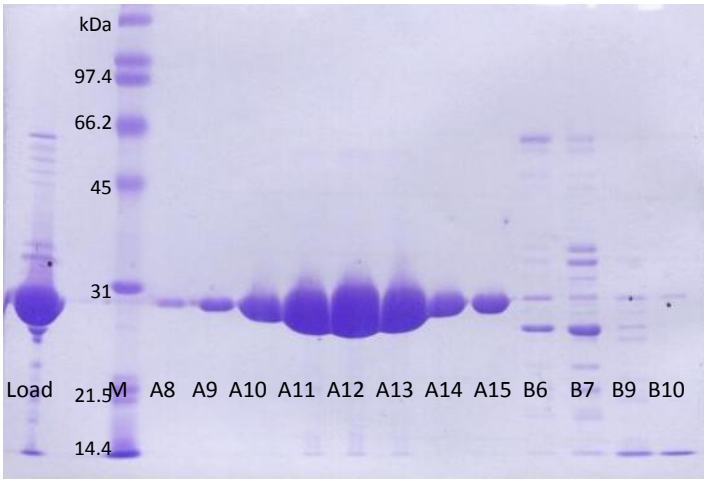

**Figure 4.8 SDS-PAGE of the E2cat anion exchange fractions**. Lanes are labelled by the fraction number shown in Figure 4.6. The load sample is shown in the lane labelled L and the standard protein markers, with their molecular weights (kDa), are shown in the lanes labelled M.

### *4.3.2.3 Gel filtration*

The anion exchange fractions containing the E2cat protein were concentrated, before a final gel filtration purification step (Figures 4.9 and 4.10). This 4 step purification method produced a pure E2cat protein sample containing no DNA/RNA in 50mM Tris-HCl, pH 8.8, and 100mM NaCl.




**4.9 Gel filtration trace of the E2 cat protein purifcation.** The $A_{280}$ readings are shown in blue, and the fraction numbers are shown along the x-axis.




**4.10 SDS-PAGE of gel filtration fractions**. The lanes are labelled by fraction number corresponding to the fractions shown in Figure 4.8; the standard protein markers are shown with their molecular weights (kDa) in the lane labelled M.

***4.3.3 1 Mass spectrometric analysis of the E2cat***

Mass spectrometric analysis was performed to ensure the purified protein was the *Tp. acidophilum* E2cat. Figure 4.11 shows 2 peaks that correspond to the predicted molecular weight of the E2cat protein with and without the start methionine: 25.455 kDa and 25.323 kDa, respectively.

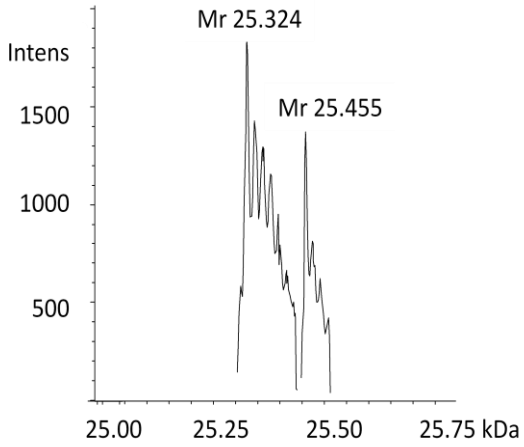



**Figure 4.11 Mass spectrometry analysis of recombinant E2cat.** The molecular weight values of the E2cat protein, with and without the start Met, are predicted to be 25.455 kDa and 25.323 kDa, respectively.

**4.3.3 E2 catalytic domain assays**

The E2cat protein was assayed to determine if the truncation altered the enzymatic activity of the protein. The E2cat, and full length non-lipoylated E2 proteins were assayed under standard conditions, the specific activity and $k_{cat}$ values of the two proteins are shown in Table 4.1.

| Enzyme | Specific activity (U/mg) | $k_{cat}$ ($min^{-1}$) |
|---|---|---|
| **E2 Catalytic domain** | 1.63 ±0.13 | 42.8 ± 3.4 |
| **Whole E2 Non-lipoylated** | 1.06 ± 0.05 | 48.9 ±2.3 |

**Table 4.1 E2cat and full-length E2 specific activity and $k_{cat}$ values.** The specific activity is shown in U/mg of E2, where 1 Unit is described as the conversion of 1 µmol of acetyl-CoA to 1 µmol of acetyl-dihydrolipoamide. The $k_{cat}$ value is the turnover number per min per E2 monomer.

The $k_{cat}$ values determined for the non-lipoylated full-length E2 and the E2cat proteins are almost identical. The removal of the flexible regions of the E2 protein has not affected enzyme activity; therefore, it would appear that this modification has not affected the structure of the E2 active site.

### 4.3.4 E2 catalytic domain thermostability

The thermostability of the E2cat protein was measured by several methods, in order to determine both the structural and functional thermostability of the protein.

#### *4.3.4.1 Dynamic light scattering*

DLS was used to assess the thermostability of the E2cat protein, by measuring the changes in the hydrodynamic diameter of the protein at increasing temperatures. The E2cat protein forms a single structure with an average diameter of ≈21nm between 25°C and 78°C. Above 80°C there is an increase in the average diameter (Figure 4.12), due to the emergence of a large protein aggregate seen by the secondary peak in Figure 4.13; however, the majority (95%) of the soluble protein remains in the smaller peak. This aggregation appears to be irreversible as, after cooling on ice and then warming back to 55°C, there is no decrease in the large aggregate peak.

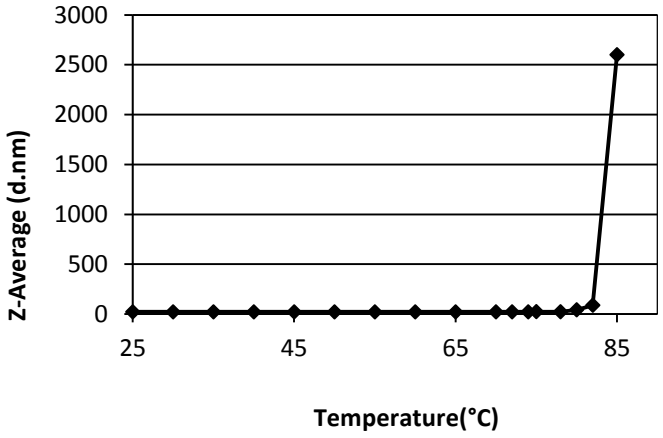



**Figure 4.12 Hydrodynamic diameter of the E2 catalytic domain with increasing temperature.**

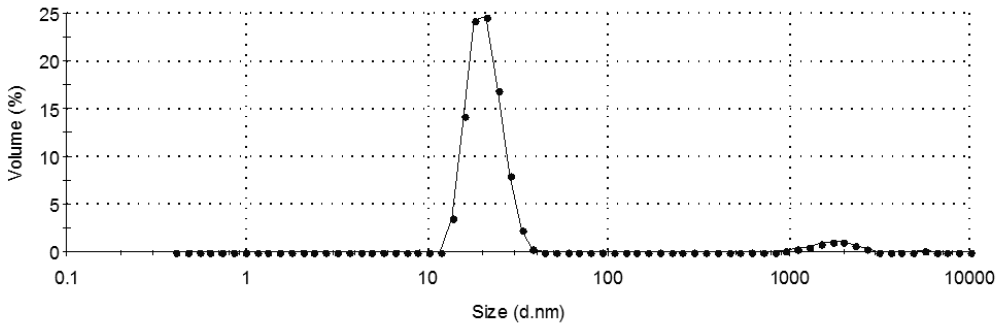



**Figure 4.13 DLS volume distribution graph at 85°C**. The majority of the signal is in the peak with a z-average ≈21nm but there is a second peak starting at 1000nm due to the presence of a large protein aggregate.

### *4.3.4.2 Heat precipitation of the E2 catalytic domain*

The thermostability of the E2cat structure was measured by comparing the ratio of soluble to insoluble protein after a 10-min incubation at various temperatures.

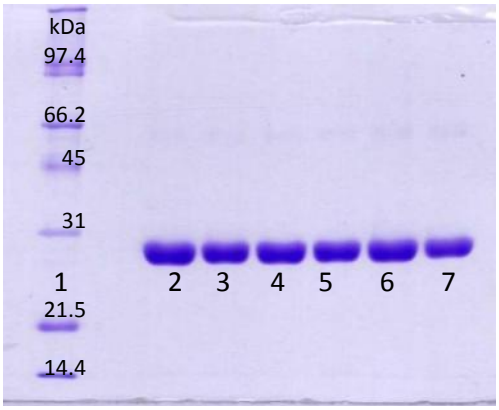



**Figure 4.14 SDS-PAGE of the soluble protein after a 10 min incubation**. Lane 1: standard protein markers, labelled with their $M_w$ (kDa); lane 2: non-heated sample; lanes 3-7: the soluble protein after 10 min at 75, 80, 85, 90 and 95°C, respectively

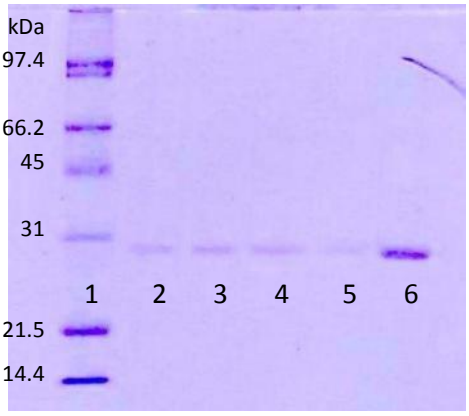



**Figure 4.15 SDS-PAGE of the insoluble protein after a 10 min incubation**. Lane 1: standard protein markers, labelled with their $M_w$ (kDa); lanes 2-6: the insoluble protein after 10 min at 75, 80, 85, 90 and 95°C, respectively.

The E2cat protein is more stable to higher temperatures than the full-length E2. Only a small quantity of E2cat begins to precipitate after 10 min at 95°C; whereas the full-length E2 begins to precipitate after 10 min at 85°C and is completely insoluble at 90°C (Figures 3.16 and 3.17). This experiment measures the temperature at which the protein precipitates, but it does not provide any information about the enzyme's ability to function after incubation at these temperatures.

### *4.3.5.3 E2cat thermal inactivation profile*

The thermal inactivation profile of the E2cat enzymatic activity was determined at 99°C (Figure 4.16). The E2cat protein is a highly thermostable enzyme that retains approximately 75% of its initial activity after a 90-minute incubation at 99°C. This is significantly more stable than the full-length E2 protein, which lost over 90% of its activity after 5 min at 90°C. The thermoactivity E2cat protein of the enzyme remains uncharacterised due to the coupled nature of the continuous E2-PTA assay.

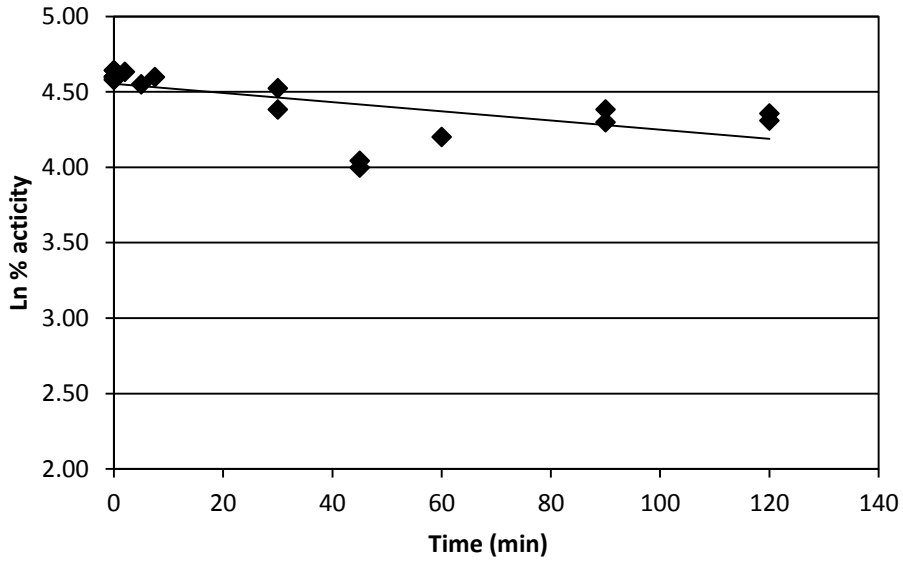



**Figure 4.16 Thermal inactivation profile of E2 catalytic domain at 99°C.** The E2cat protein was heated in 50mM Tris-HCl, pH8.8, and 100mM NaCl.

## 4.4 Conclusion

The C-terminal catalytic domain of the *Tp. acidophilum* E2 enzyme has been successfully cloned and expressed. The E2cat protein assembles into a large structure which retains acetyl-transferase activity with a similar $k_{cat}$ value to the full-length E2 protein; this suggests that the structure and assembly of the E2cat domain has not been affected by the removal of the flexible linker regions. The E2cat protein is significantly more thermostable than the full-length E2, and this result raises the question as to why this domain is hyper-thermostable whereas the whole E2 protein is not. One possible explanation stems from the observation that thermostable enzymes tend to have shorter loops, and reduced N- and C-terminal extensions, as these flexible structures can be sites of thermal denaturation (Thompson and Eisenberg, 1999). In the case of the E2 protein, the lipoyl domain must be able to move freely between the active sites of the E1 and E3 enzymes; consequently, the E2 catalytic domain may be required to be hyper-thermostable, to resist the thermal denaturation forces resulting from the movement of the flexible regions of the protein.

# Chapter 5: The structure of the *Tp. acidophilum* dihydrolipoamide transacetylase

## 5.1 Introduction

### 5.1.1 Determining the structure of the *Tp. acidophilum* E2 catalytic domain

The previous chapter described the cloning and expression of the *Tp. acidophilum* E2 catalytic domain; the aim of this chapter was to determine the structure of this hyperthermostable protein, in order to begin to understand the structural basis of the thermostability of the OADHC. The E2 catalytic domain was cloned and expressed, without its flexible regions, with the aim of crystallising the protein and generating a high-resolution X-ray crystallographic structure. This chapter also covers other biophysical techniques including small-angle X-ray scattering, sedimentation equilibrium analytical ultracentrifugation, and transmission electron microscopy; these techniques were used to validate the E2cat crystal structure, and to compare it with the whole E2, to establish if removing the flexible regions alters its structure and/or assembly.

The E2 structures currently available are either cubic 24-mers or dodecahedral 60-mers; in all cases, the E2 polypeptide chains form trimers that then assemble into a large hollow cage, composed of either square or pentagonal faces, with a large opening in the centre of each face (Figure 5.1). The E2 transacylase active site is located in a channel between the monomers of each trimer; this channel extends from the outside to the inside of the assembled E2 structure. The lipoic acid moiety with the acyl group attached enters the active site from the outside and the CoA enters from the inside; once the acyl-groups has been transferred, the acyl-CoA leaves the active site and diffuses through the windows in each face, and the lipoyl domain then visits an E3 in order to be re-oxidised and thus regenerated (Mattevi et al., 1992).

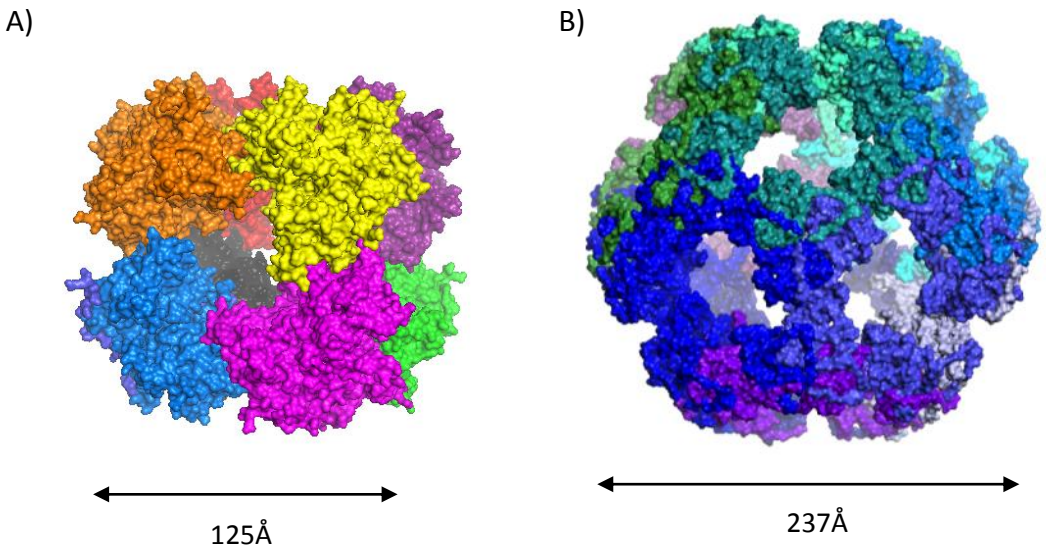



**Figure 5.1 Surface representations of A) the cubic 24-mer and B) the dodecahedral 60-mer.** The cubic 24-mer is the pyruvate dehydrogenase complex from *Azotobacter vinelandii* (PDB code: 1EAF) (Mattevi et al., 1992) and the 60-mer is from the pyruvate dehydrogenase complex from *Geobacillus stearothermophilus* (PDB code: 1B5S) (Izard et al., 1999). The maximum diameter is shown below each structure.

### 5.1.2 X-ray crystallography

X-ray crystallography is a powerful technique capable of determining high-resolution protein structures. The first hurdle to overcome is growing a good quality crystal from the required protein, which involves producing an ordered, three dimensional arrangement of the protein molecules held together by non-covalent interactions. Each protein crystal comprises a lattice of unit cells in which several protein molecules may be packed. Slow, controlled precipitation of a protein is required to produce a good quality protein crystal containing identical unit cells, which are packed into a regular lattice. The exact conditions (buffer, pH, precipitant type and concentration, presence of additional ions or sugars and protein concentration) required to produce a crystal rather than an amorphous precipitate vary depending on each protein (Rhodes, 1993).

Structural information about the protein is determined by mounting the crystal in an X-ray beam and recording the pattern of diffracted X-rays. Each exposure of the crystal to the X-ray beam produces a scattering pattern containing two-dimensional information about the crystal; the crystal is rotated in the beam to provide multiple images that are compiled to construct a three-dimensional image. The quality of packing within the crystal affects the resolution of the data recorded - the more ordered and similar each of the unit cells, the higher the resolution (Rhodes, 1993).

An electron density map can be created from the diffraction pattern using the Fourier equation; this process is simplified if another similar structure is available, as molecular replacement can solve the phase problem by predicting the X-ray diffraction pattern created by another similar protein, and comparing it with the experimental data (Rhodes, 1993). The level of detail contained within the electron density map depends on the quality of data collected - the higher the resolution the more accurately each atom can be positioned. The protein structure is modelled into the electron density map and the positioning of each atom improved with iterative rounds of refinement.

Determination of a high-resolution protein structure by X-ray crystallography requires the protein to be packed into a crystalline state, a process that is not possible for all proteins or protein regions. Flexible proteins can adopt multiple conformations, and therefore do not pack into ordered crystals; the full length E2 protein is an example of a flexible protein which to date has not been crystallised. Another limitation that should be considered is the distortion of protein structures or alteration of interaction caused as a result of crystal packing; the structures determined by X-ray crystallography may, in some cases, differ from the protein structure or arrangement in solution (Rhodes, 1993).

### 5.1.3 Small-angle X-ray scattering

Small-angle X-ray scattering (SAXS) is a method used to determine the molecular weight, oligomeric state, and shape of a protein in solution. The X-ray radiation is scattered as it passes through the solution containing protein and/or buffer (Figure 5.2). The scattering pattern produced is a rotational average, as the protein will be in multiple orientations within the solution at any one time, unlike the ordered crystal packing seen in X-ray crystallography; this pattern is recorded on a detector up to an angle typically of 3°. The scattering pattern of a protein is measured at several different concentrations, to establish if there are any concentration-dependent changes in the structure or assembly. Each sample is also subjected to multiple exposures to detect any radiation damage. The 'buffer-only' scattering pattern is deducted from the protein pattern and the data are normalised by comparison with BSA standards (Figure 5.3). The normalised scattering intensity of the protein can then be plotted versus the scattering vector (Figure 5.3B), the shape of this plot supplies information about the protein's size and overall shape. Examples of the characteristic scattering patterns produced by a variety of shapes are shown below in Figure 5.4 (Jacques and Trewhella, 2010).

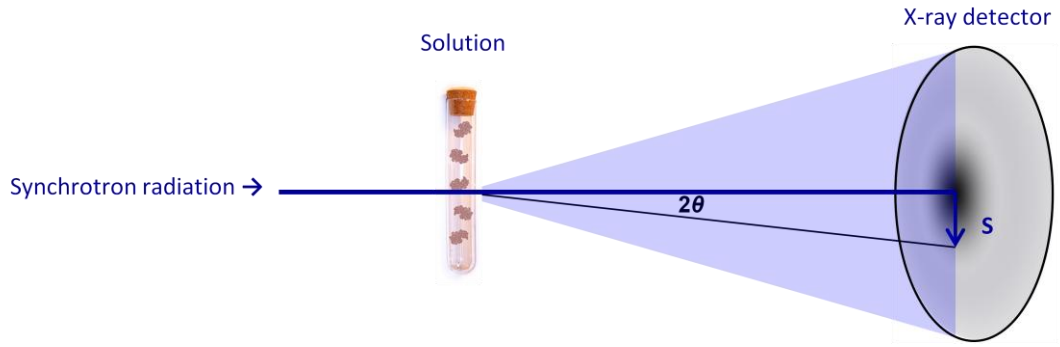



**Figure 5.2 Experimental arrangement to collect small-angle X-ray scattering data.** The synchrotron X-ray radiation passes through the solution containing protein and buffer. The protein and buffer molecules scatter the X-rays, producing a rotation averaged scattering pattern that is collected usually up to an angle of 3°. The scattering intensity versus the scattering vector (S) is recorded and analysed to provide low-resolution structural information about the protein.

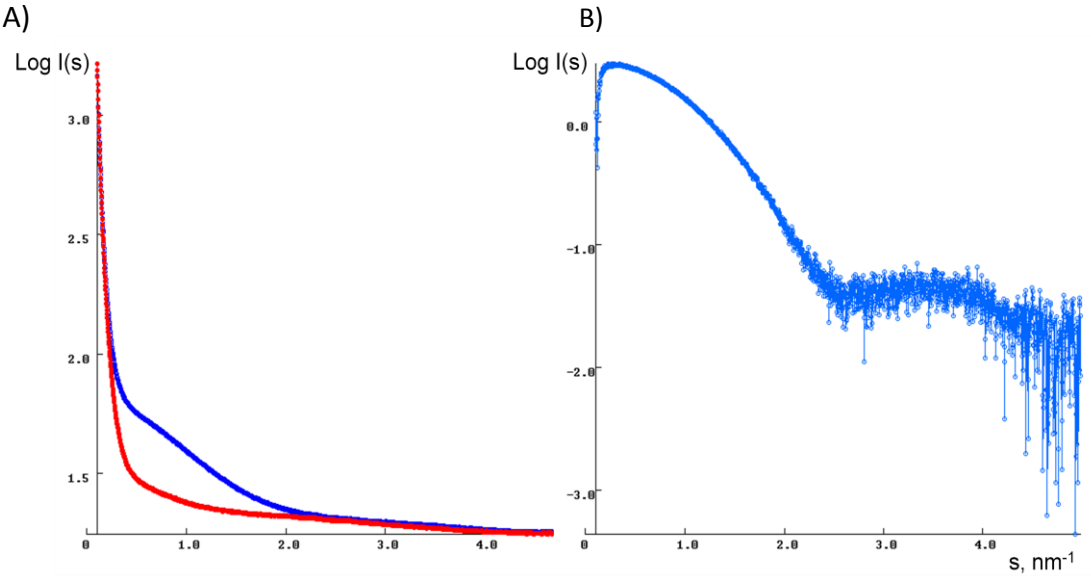



**Figure 5.3 Examples of scattering intensity graphs of the logarithm of the intensity (Log I(s)) versus the scattering vector (s, nm).** Graph A shows the scattering pattern of a protein sample in blue and a 'buffer-only' control in red. Graph B is the normalised scattering pattern of the protein minus the buffer background.

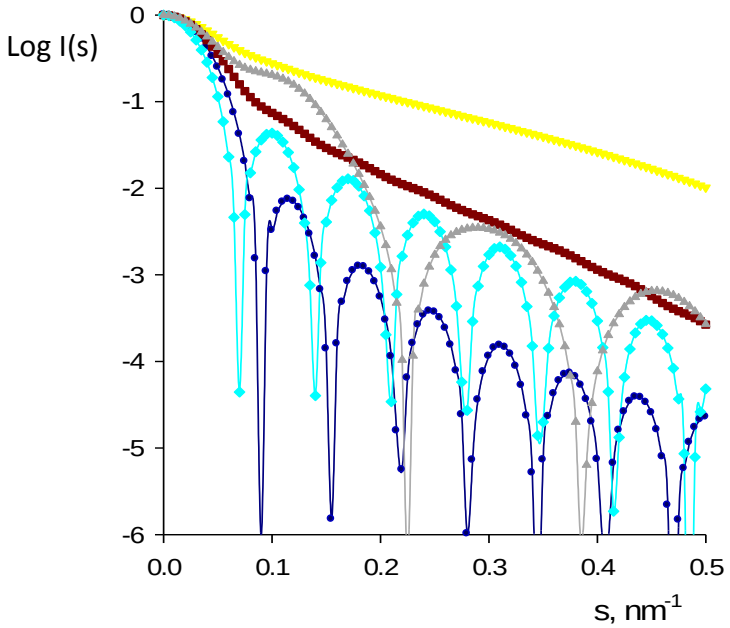



**Figure 5.4 Examples of characteristic scattering patterns produced by a range of theoretical shapes.** The coloured lines represent typical scattering patterns produced by following shape molecules: a rod (yellow), dumbbell (grey), flat disk (brown), hollow sphere (cyan) and solid sphere (dark blue).

The scattering pattern can then be transformed to produce a series of graphs that provide the various parameters used to determine the size and shape of the protein, the presence of any aggregation, and the quality of the data (Jacques and Trewhella, 2010). The secondary plots and the information provided are detailed below:

- Guinier plot – radius of gyration (*R*g), molecular weight and assessment of the data quality
- Porod plot – particle volume
- *P*(r) plot – maximum linear dimension ($D_{max}$), and identification of elongated protein structures

An *ab initio* bead model can then be created by predicting the scattering pattern produced by a hollow sphere and comparing this to the experimental data; the hollow sphere is then modified and the scattering pattern determined again. After multiple iterations of this process, the aim is to produce a final model that generates a predicted scattering pattern that is the same as the experimental scattering data. A higher-resolution crystallographic structure of the protein, if available, can be modelled into the *ab initio* bead model to confirm statistically if the crystallographic and SAXS structures are consistent with one another.

## 5.2 Methods

### 5.2.1. Protein Expression and purification

#### *5.2.1.1 E2 and E2cat expression*

The E2 and E2cat proteins were expressed following the method described in Chapter 3.2.1.

#### *5.2.1.2 E2 and E2cat protein purification*

The E2 and E2cat proteins were both purified by the method described in Chapter 4.2.2, resulting in protein samples which were >95% pure in 50mM Tris-HCl, pH 8.8, and 100mM NaCl

### 5.2.2 Transmission electron microscopy

Purified E2cat protein (25µg/ml) was attached to Formvar carbon-coated grids, negatively-stained and visualised by Ursula Potter (University of Bath). The carbon-coated grids were prepared by overnight exposure to UV-light. The E2cat protein (25µg/ml in 50mM Tris-HCl, pH 8.8, 100mM NaCl) was attached to the carbon-coated grids by adding 30µl of protein and incubating for 1 min, any unattached protein was then removed by two washes with 50µl of $dH_2O$. The protein coated grids were blotted dry and stained with 50µl of either 2% (w/v) uranyl acetate pH 4, or 2% (w/v) phosphotungstic acid pH 8; after 30 sec any excess stain was removed and the grids air-dried before visualisation on a Jeol JEM1200EXII (Tokyo, Japan).

### 5.2.3 Determination of the X-ray crystallographic structure of the E2 catalytic domain

#### *5.2.3.1 Crystal condition screens*

An Art Robbins Nano dispenser robot was used to set up PACT and JCSG 96-well screens (Molecular Dimension, Suffolk, U.K.). Each condition had 3 sitting drops with a differing volume ratio of protein (~10mg/ml) to well solution: 1:2 1:1 and 2:1, totalling 300nl. The plates were sealed with a plastic sheet and stored at 16°C. The screens were checked every two days for two weeks, and then weekly for several months.

#### *5.2.3.2 Optimisation*

Conditions from the 96-well plates that resulted in protein crystals or granular precipitation were selected for optimisation in a 24-well hanging-drop plate. The conditions identified from the 96-well screens were initially replicated in hanging drops containing 1.5µl:1.5µl well solution to protein. The conditions were then varied to improve the size and quality of the crystals, including changing the buffer composition and pH, precipitant and its concentration, protein concentration and the addition of other metal ions and sugars.

#### *5.2.3.3 X-ray diffraction data collection*

The crystals were removed from the drops with a cyroloop and soaked in 0.1M Na acetate, pH 5.5, 15% (v/v) 2-methyl-2,4-pentanediol (MPD), 0.20M NaCl and 10% (v/v) ethylene glycol, before flash-freezing in liquid nitrogen. X-ray diffraction data were collected at the Diamond LightSource (Oxfordshire, U.K.) on an ADSC Q315 3x3 CCD detector on station I04 (at wavelength $\lambda$=0.9702 Å) by Dr Jean van den Elsen and Dr. Susan Crennell (University of Bath, U.K.). The data were then processed by Dr Susan Crennell using the HKL2000 package.

#### *5.2.3.4 Model building and refinement*

Molecular replacement was carried out by Dr. Susan Crennell with BALBES (Long et al., 2007) using the *Azotobacter vinelandii* dihydrolipoamide acyltransferase structure (PDB access code 1EAF). The model was built and refined using COOT (Emsley and Cowtan, 2004) and REFMAC5 - CCP4i software (Bailey, 1994). The protein structure was deposited in the Protein Data Bank (PDB access code: 3RQC). Images were created using PyMOL.

### 5.2.4 Small-angle X-ray scattering

The E2cat protein was purified as described in Chapter 4.3.2. The E2cat sample was analysed in two buffers: (1) 50mM sodium phosphate, pH 7.0, and 100mM NaCl; (2) 50mM Tris-HCl, pH 8.8, and 100mM NaCl. The SAXS data were collected at the X33 beam line, EMBL Hamburg Outstation, (DORIS-III storage ring at Deutsches Elektronen-Synchrotron (Roessle et al., 2007)) at 10°C using an automatic sample changer (Round et al., 2008); data were collected for four protein concentrations: 1.0, 1.5, 2.0 and 2.5mg/ml. Two successive two-minute exposures of the 1.0 mg/ml sample detected no radiation damage. The data were processed and analysed by Dr. Dmitri Svergun (EMBL Germany),

Dr. Peter Konarev (EMBL, Germany) and Dr Jean van den Elsen (University of Bath, U.K.). The results were analysed using standard procedures; the data were extrapolated to zero solute concentration using PRIMUS (Konarev et al., 2003), GNOM was used to determine forward scattering and the radius of gyration, that were used to calculate the proteins maximum dimensions and intraparticle distance (Svergun, 1992). Ten low-resolution *ab initio* bead models of the E2 catalytic domain were produced by the DAMMIFF programme (Franke and Svergun, 2009, Svergun, 1999) and averaged using the DAMAVER programme (Volkov and Svergun, 2003). The high-resolution scattering models were calculated with the program CRYSOL (Svergun et al., 1995), and the rigid-body modelling of the E2cat was performed using the program SASREF without symmetry restrictions (Petoukhov and Svergun, 2005). The 4Å X-ray crystallographic structure previously determined was then modelled into the *ab initio* bead model.

### 5.2.5 Sedimentation equilibrium analytical ultracentrifugation

All experiments were performed and results analysed by Dr. Jacqui Marshall (University of Bristol), using an An60Ti rotor in a Beckman XL-A ultracentrifuge at 20°C. A 6-sector centrepiece centrifuge cell was loaded with 120µl of three different protein concentrations, along with three buffer-only controls, to give absorbance values in each 1.2 cm cell of approximately 0.8, 0.5 and 0.3. The different concentrations were analysed to detect any concentration-dependent changes in the protein.

The samples were centrifuged at various speeds, and the absorbance values were measured every 6 h until equilibrium was reached. The buffer background was measured at the end of the experiment after a 6 h clearing run at 40,000 rpm.

The data were fitted using ORIGIN to the model for a single ideal protein species, based on the Svedberg equation (shown below). The partial specific volume ($\overline{V}$) of the protein and buffer density ($\rho_o$) were estimated from the protein sequence and the buffer composition, respectively, using SEDNTERP:

[http://www.jphilo.mailway.com/download.htm]

The predicted molecular weight ($M_w$) and absorbance coefficient of each protein was determined using the Expasy Protparam (Gasteiger et al., 2003).

Svedberg equation:

$$M_W = \frac{RTs}{D(1 - \overline{V}\rho_o)}$$

Where $M_w$ = molecular weight (Da) of the protein assembly, R = universal gas constant (8.314 x $10^3$ J/K/Kg mol mass), T = temperature (K), s = sedimentation coefficient (sec), D = diffusion coefficient ($m^2$/sec), $\overline{V}$ = partially specific volume of the protein (ml/g), and $\rho_o$ = density of the buffer (g/ml).

## 5.3 Results and discussion

### 5.3.1 Transmission electron microscopy

Transmission electron microscopy was used as an initial low-resolution method to determine if the E2cat protein assembled into a large hollow structure. Figure 5.5 shows a hollow protein structure approximately 17nm in diameter; however, it is not possible to determine the structure from this image as the symmetry is not clear.

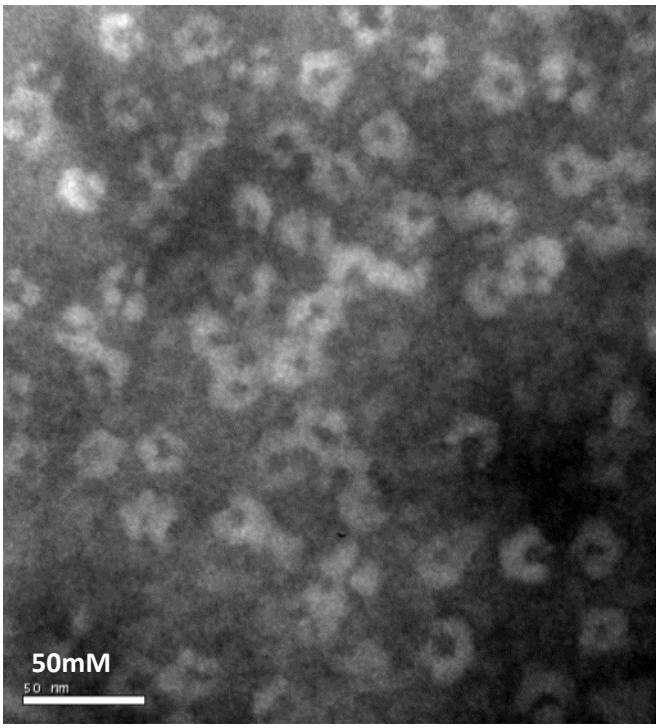



**Figure 5.5 Transmission electron microscopy image of the E2cat protein.** E2cat protein visualised by negative staining with 2% (w/v) phosphotungstic acid, pH 8.0

### 5.3.2 E2catalytic domain crystal trials

#### *5.3.2.1 Initial 96 well plate crystal screen*

PACT and JCSG 96-well screens (Molecular Dimensions, Suffolk, U.K.) were used to identify conditions that resulted in crystals or granular precipitate. The PACT screen resulted in precipitate in all 96 conditions and all concentrations. The JCSG screen identified three conditions (Table 5.1) that were optimised in 24-well plates.

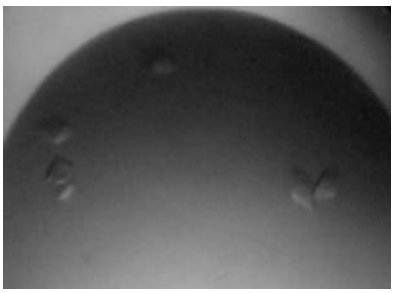
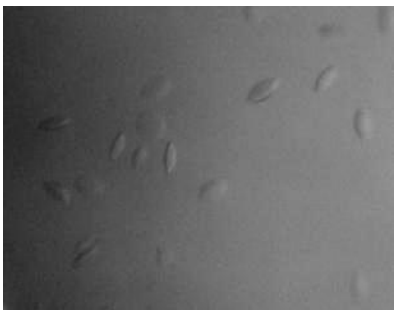
| Condition | Crystals |
|---|---|
| 0.1M Tris-HCl, pH 8.5<br>1.26M Ammonium sulphate<br>0.2M lithium sulphate |  |
| 0.1M Caps, pH 10.5<br>40% MPD |  |
| 0.1M Hepes, pH 7.5<br>20% Jeffamine M-600 |  |

**Table 5.1 Examples of the crystals that grew in three different conditions within the JCSG 96-well crystal screen.** The images show the crystal forms that grew in each condition within 1 month.

#### *5.3.2.1 Crystal optimisation*

The 0.1M Caps, pH 10.5, 40% MPD condition was optimised for protein concentration, pH, buffer, precipitation and additional salts. The final condition that produced crystals suitable for X-ray diffraction contained 1.5µl of 10mg/ml protein, mixed with 1.5µl of well solution containing 0.1M Na acetate, pH 5.5, 12% MPD, and 0.2M NaCl, and storing at 16°C for approximately 2 weeks (Figure 5.6). All attempts to reproduce and optimise the other two conditions were unsuccessful in terms of crystal formation.

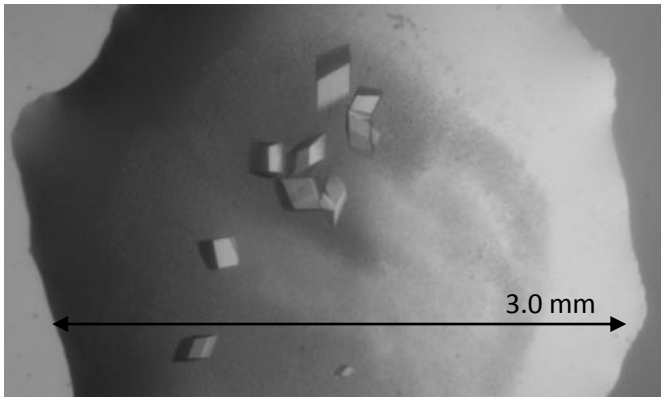



**Figure 5.6 E2cat crystal in 0.1M Na Acetate, pH 5.5, 12% MPD, 0.20M NaCl.** Crystals grew over two weeks at 16°C.

### 5.3.3 The structure of the *Thermoplasma acidophilum* E2 catalytic domain

#### *5.3.3.1 X-ray diffraction data collection*

Diffraction to 4Å was achieved with a diamond-shaped crystal that was flash frozen in 0.1 M Na acetate, pH 5.5, 15% MPD, 0.20 M NaCl, with an additional 10% ethylene glycol. A variety of crystal sizes, morphologies and cryoprotectants were tried but no further improvements were made to the diffraction data.

#### *5.3.3.2 E2cat model building and refinement*

The X-ray diffraction data were collected at the Diamond Light Source and processed to 4.0 Å (Table 5.2); the refinement statistics are given in Table 5.3.

| | | |
|---|---|---|
| **Space Group** | | R32 (H32) |
| **Cell parameters** | a=b, c | 206.3Å, 438.4Å |
| | α=β, γ | 90.00°, 120° |
| **Resolution (Å) (last shell)** | | 4.00 (4.14 - 4.00) |
| **Number of observations** | | 670508 |
| **Number of unique reflections** | | 29696 |
| **Completeness (last shell)** | | 98.0 (99.3) |
| **R merge (%)(last shell)** | | 12.7 (48.5) |
| **Average I/σI (last shell)** | | 7.9 (1.58) |
| **Redundancy** | | 3.7 (3.5) |

**Table 5.2 X-ray data collection statistics from the E2cat crystal**

| | | |
|---|---|---|
| **Resolution range** | | 44.95 - 4.00Å |
| **Number of protein atoms** | | 12157 |
| **Number of solvent atoms** | | None |
| **Average B value** | | 97.108Å$^2$ |
| **R-crystal** | | 25.3% |
| **R-free** | | 32.8% |
| **Rms deviations** | **Bonds** | 0.015Å |
| | **Angles** | 1.806° |

**Table 5.3 E2cat structure refinement statistics**

The structure of the *Tp. acidophilum* E2cat protein was solved by molecular replacement using the *A. vinelandii* dihydrolipoamide acyltransferase structure (PDB access code 1EAF). This produced a model comprising 6 monomers arranged as two adjacent trimers, with clear electron density for a seventh monomer. An additional single monomer was modelled into this density resulting in a 7 chain asymmetric unit (Figure 5.9). The polypeptide backbone of this seven-chain model (Figures 5.7 - 5.9) was refined to give final values of $R_{work}$ 25.3% and $R_{free}$ 32.8% (Table 5.3).

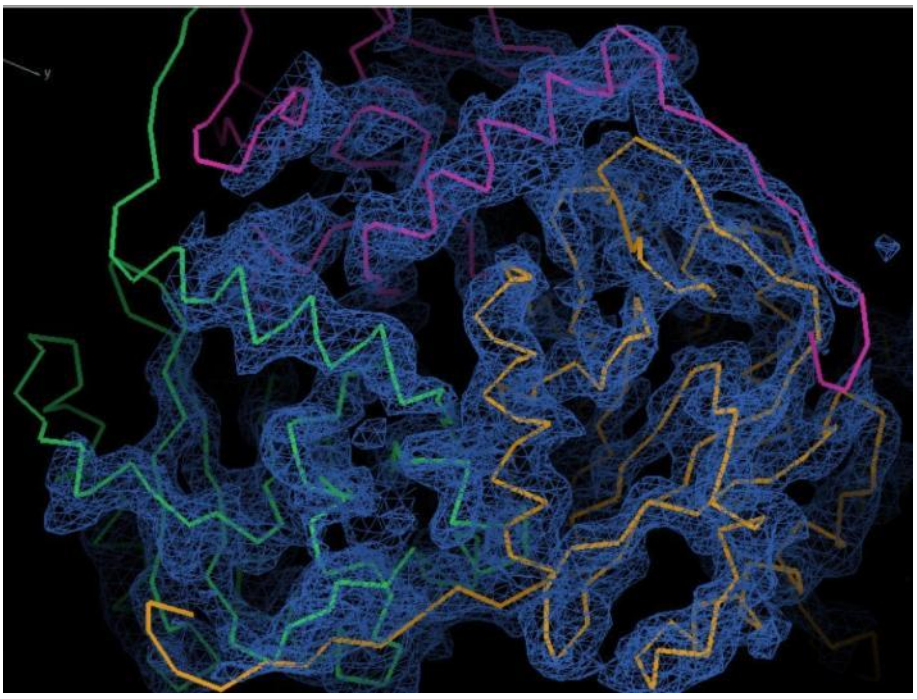


**Figure 5.7 The E2cat trimer polypeptide backbone and the electron density map.**

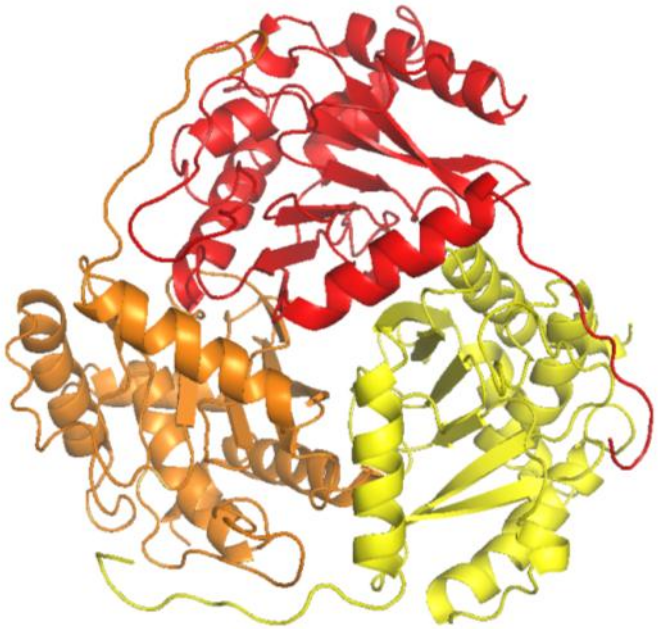


**Figure 5.8 A ribbon diagram of the *Tp. acidophilum* E2cat trimer polypeptide backbone structure.**

**Figure 5.9 Ribbon diagram of the seven chains comprising the E2cat asymmetric unit.** The seventh chain (black) sits on a three-fold axis of symmetry; the other 2 monomers in the trimer are created by symmetry.

#### *5.3.3.1 The 42-mer structure of the Thermoplasma acidophilum E2*

The 7-chain asymmetric unit comprises two trimers and one monomer; the seventh monomer creates a third trimer as it is located on a three-fold symmetry axis; the complete E2cat structure was created by positioning the other symmetry related trimers (Figure 5.10). The resulting structure was neither the expected cubic 24-mer (432-symmetry) nor the dodecahedral 60-mer (532-symmetry); instead, the *Tp. acidophilum* E2cat assembled into a novel 42-mer, with 32-symmetry (Figure 5.10) (Marrott et al., 2012). The 1.08 MDa, 42-mer structure is an oblate sphere, comprising three square and six pentagonal faces, with a maximum diameter of 220 Å and a minimum of 190 Å. The large openings located in the centre of the square and pentagonal faces have a diameter of 40 Å and 52 Å respectively. The cubic *A. vinelandii* E2cat has an overall diameter of 120Å, with 30 Å wide windows (Mattevi et al., 1991b); and the dodecahedral *G. stearothermophilus* E2cat measures 237 Å in diameter, with 52 Å windows (Izard et al., 1999). The *Tp. acidophilum* 42-mer dimensions fit between these two previously described structures. The windows in the pentagonal faces are similar to those of the dodecahedral structure; however, the windows in the *Tp. acidophilum* square faces are larger than those in the cubic 24-mer structure. The *Tp. acidophilum* structure has a high solvent to protein ratio of 76%, which may have been responsible for limiting the data collection resolution to 4Å; this is similar to the 4.4Å resolution achieved by Izard et al (1999) for the *G. stearothermophilus* E2 which has an even higher solvent content of 89%.

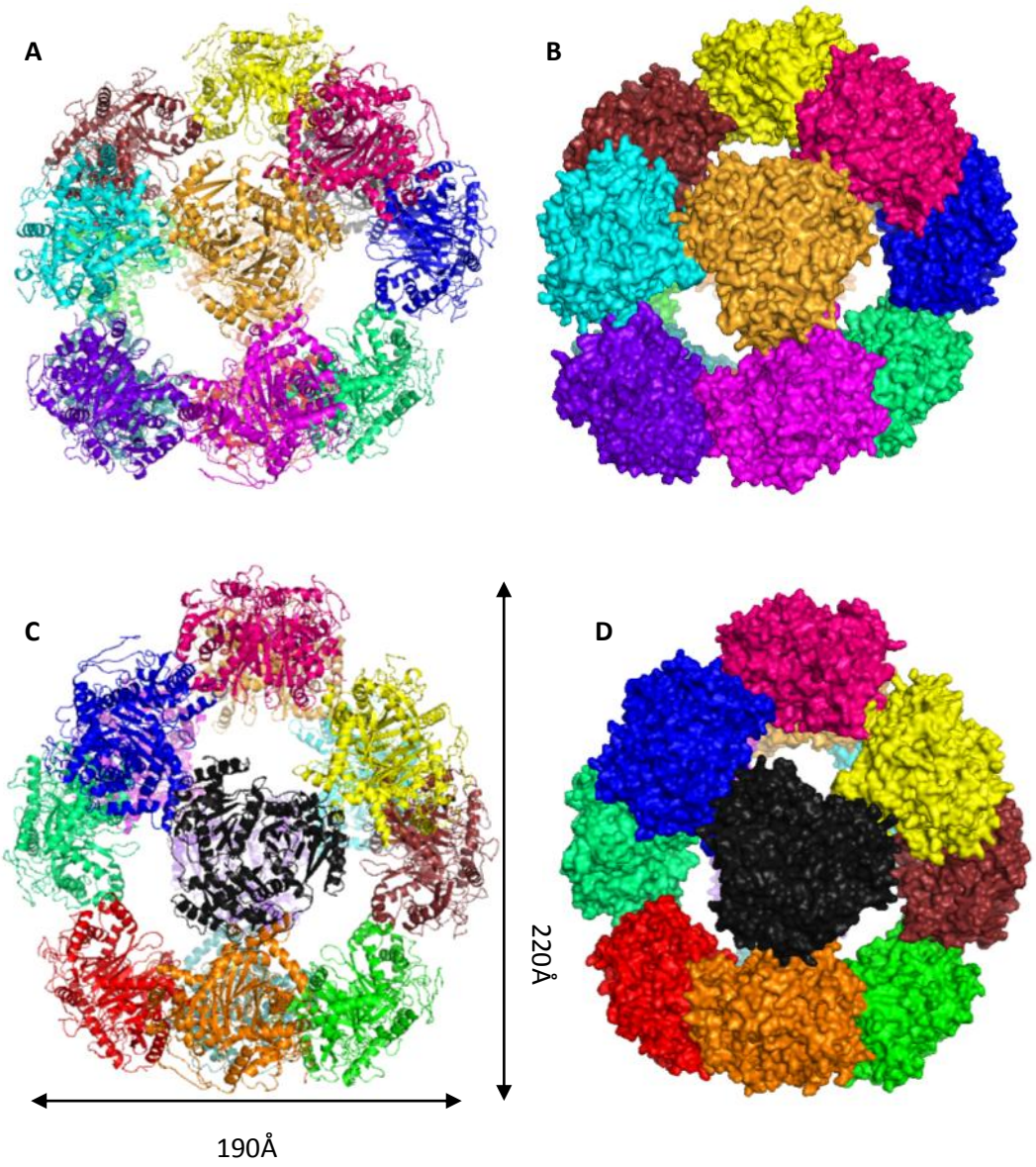



**Figure 5.10 The assembled 42-mer *Tp. acidophilum* E2cat structure.** The ribbon diagrams shown in figures A and C illustrate the backbone structure of the E2cat from 2 different angles, showing two pentagonal and one square face. B and C show a space-filling representation of A and C respectively. The three polypeptide chains within each trimer are coloured the same. PDB accession code: 3RQC.

### 5.3.4 Analysis of the E2 catalytic domain structure

#### *5.3.4.1 Sequence and structural alignments with other E2 catalytic domain structures*

The polypeptide backbone structure can be determined utilising the 4Å resolution data but not the position of the side chains. In order to analyse the structure further, comparisons were performed with other higher-resolution E2cat structures in which the side-chain positions have be determined, including: the *A. vinelandii* PDHC (2.6Å) (Mattevi et al., 1991b), the bovine branched-chain OADHC (2.7Å) (Kato et al., 2006), the *E. coli* 2-oxoglutarate dehydrogenase complex (OGDHC) (3.0Å) (Knapp et al., 1998) and also the lower resolution *G. stearothermophilus* PDHC (1B5S) (4.4Å). The alignment of the *Tp. acidophilum* E2cat sequence with the sequences from known E2cat crystal structures are shown in Figure 5.11 with the key residues involved in trimer-trimer interaction and catalytic activity highlighted.

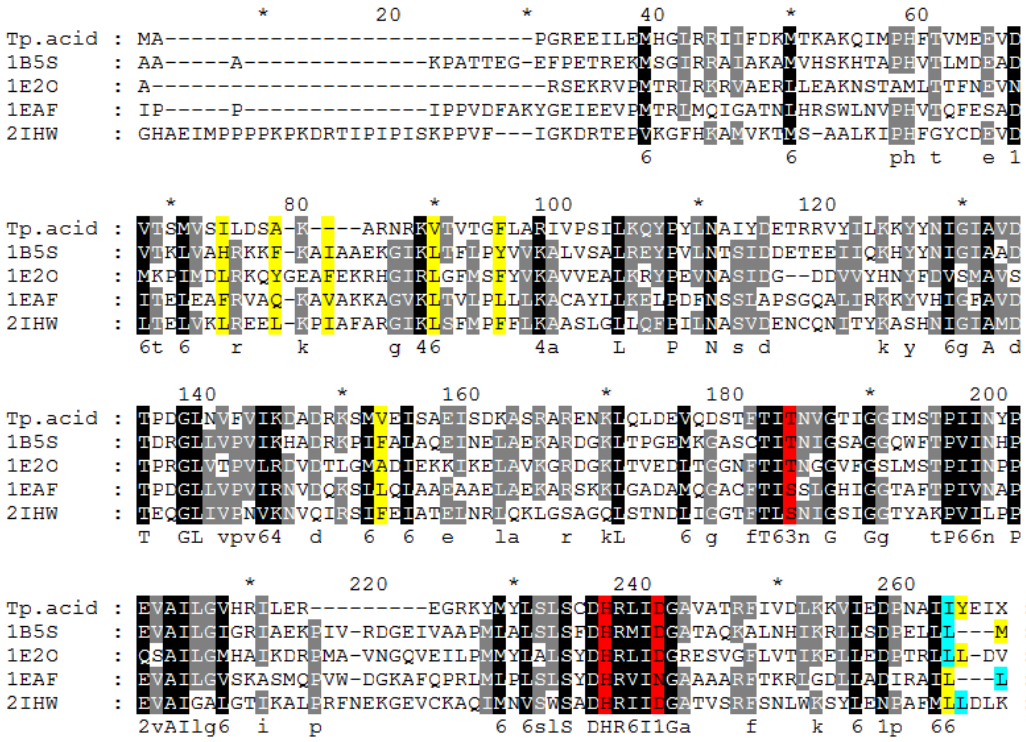


**Figure 5.11 *Thermoplasma acidophilum* E2cat sequence alignment with sequences from other known E2cat structures.** Tp.acid is the *Tp. acidophilum* sequence E2 catalytic domain. The other E2cat sequences are labelled with their PDB codes: 1B5S: *G. stearothermophilus* PDHC; 1E2O: *E. coli* OGDHC, 1EAF: *A. vinelandii,* PDHC; 2IHW: bovine branched-chain OADHC. The three key active site residues are in red, the anchor resides are highlighted in cyan, and the hydrophobic pocket residues yellow in yellow.

### *5.3.4.2 Active site residues*

The active site residues of the E2 transacylase are positioned within a tunnel which allows entry of the lipoic acid from the outside and the CoA from the inside of the hollow E2 core (Mattevi et al., 1992). The key active site residues in the *Tp. acidophilum* E2cat protein can be predicted based on alignments with other E2cat sequences; the active site is found between 2 monomers with one chain contributing a threonine (Thr 151) residue and the other providing the histidine (His 195) and aspartic acid (Asp 199) (Figure. 5.12). The active site side chains appear to be in the correct position in the *Tp. acidophilum* E2cat structure, suggesting that the polypeptide backbone has been positioned correctly, as although the exact side chain position is not known they are orientated correctly on the α-helices and β-sheets.

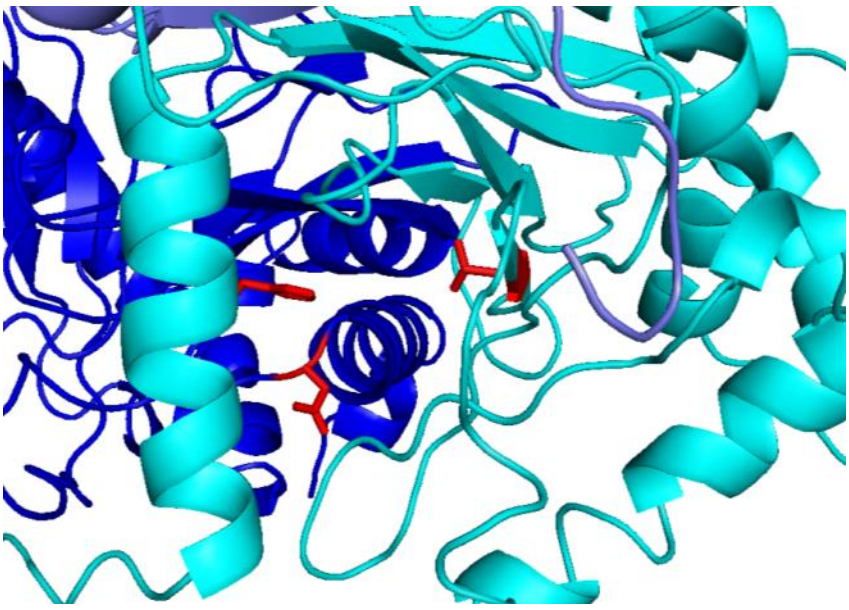


**Figure 5.12 The E2cat active site shown in channel between 2 monomers (cyan and blue).** The blue monomer contributes the Asp (199) and His (195), and the Thr (151) is provided by the cyan monomer.

### *5.3.4.4 Predicting the anchor residue key to inter-trimer interactions*

An anchor residue is predicted to be important in the formation of inter-trimer interactions, in both the 24- and 60-mer structures; it is a branched side-chain residue that extends from one trimer into a hydrophobic pocket on the second trimer and *vice versa* (Izard et al., 1999). Based on the sequence alignments of the *Tp. acidophilum* sequence along with other known structure sequences (Figure 5.11), two possible isoleucine anchor residues (E2cat Ile 220 and Ile 221) have been identified. Comparing the locations of these two residues on the crystal structure, isoleucine 221 appears the more likely candidate, as it is positioned in the correct orientation on the $3_{10}$ helix (Figure 5.14). There is electron density for the Ile 221 side chain extending from one trimer towards the adjacent trimer; however, due to the 4Å resolution the precise position of the side chain cannot be confirmed (Figure 5.13). The consequences of mutating and removing this anchor residue are detailed in Chapter 6. The trimers located on both the square and pentagonal faces interact via the same putative Ile 221 anchor residue; this ball and socket-like interaction may allow the flexibility required for the trimer to form the two different faces.

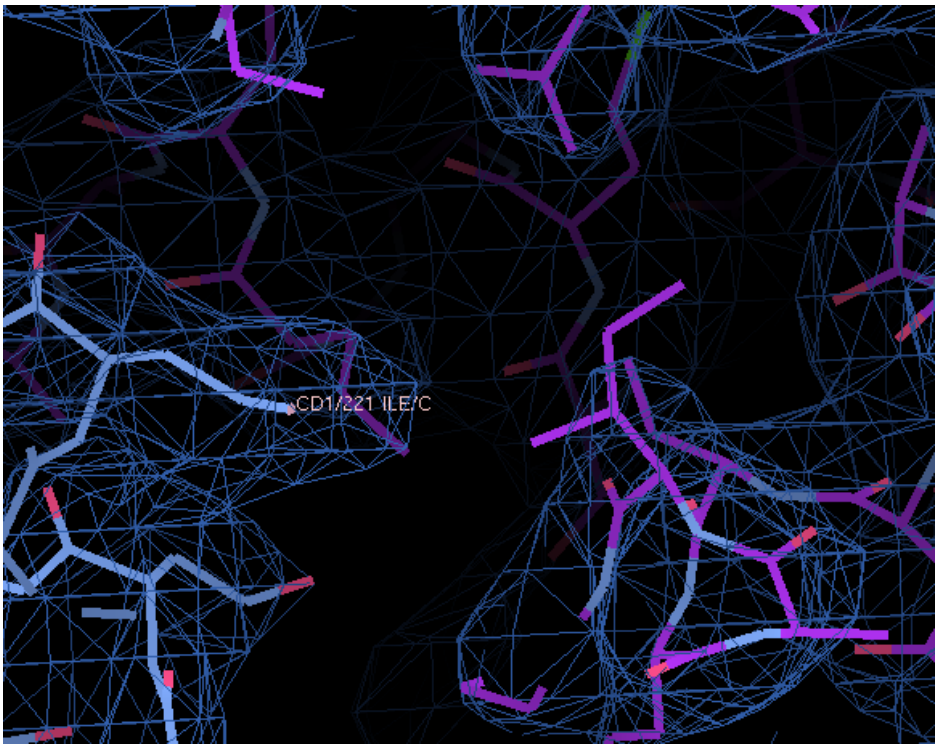
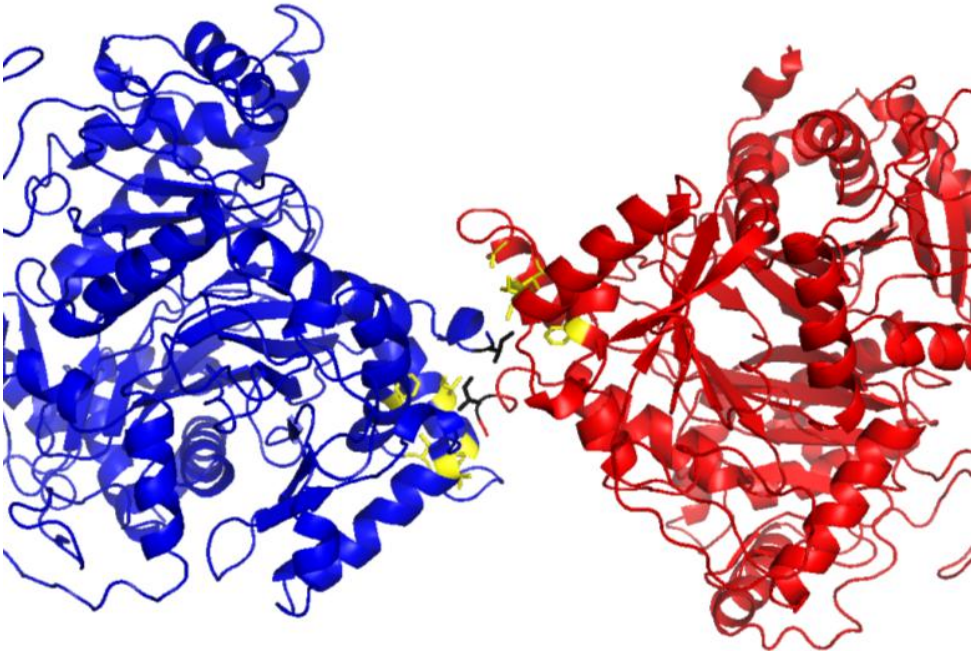



**Figure 5.13 The E2cat backbone structure with the side-chain of the possible anchor residue (isoleucine 221) with the electron density superimposed.** The isoleucine 221 side chain had been highlighted with the label CD1/221 ILE/C.

A)



B)

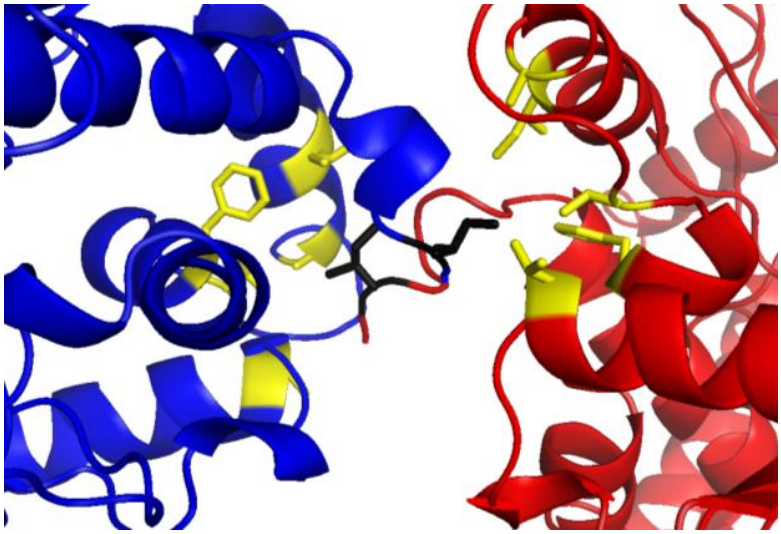


**Figure 5.14 Crystal structure highlighting the possible anchor residues involved in inter-trimer interactions**. The ribbon diagram backbones of the two adjacent trimers are shown in red and blue. The Ile 221 side-chains are shown in black, extending from each trimer into a yellow hydrophobic pocket on the adjacent trimer.

### *5.3.4.5 Formation of square and pentagonal faces within the 42-mer*

All current E2 structures comprise trimers that associate into a hollow cage with either square or pentagonal faces. The 24-, 42-, and 60-mer arrangements are not as a result of differences in the trimer structure (Figure 5.15), rather the angle at which the trimers interact (Figure 5.16). Unlike the other known structures, the *Tp. acidophilum* trimers appear to be capable of interacting differently with one another in order to form both square and pentagonal faces (Figure 5.17).

**Figure 5.15 Overlaid E2 trimer structures of *Tp. acidophilum* 42-mer, a cubic 24-mer and a dodecahedral 60-mer**. The *Tp. acidophilum* ribbon diagram is shown in black, the *A. vinelandii* is shown in orange and the *G. stearothermophilus* in blue.

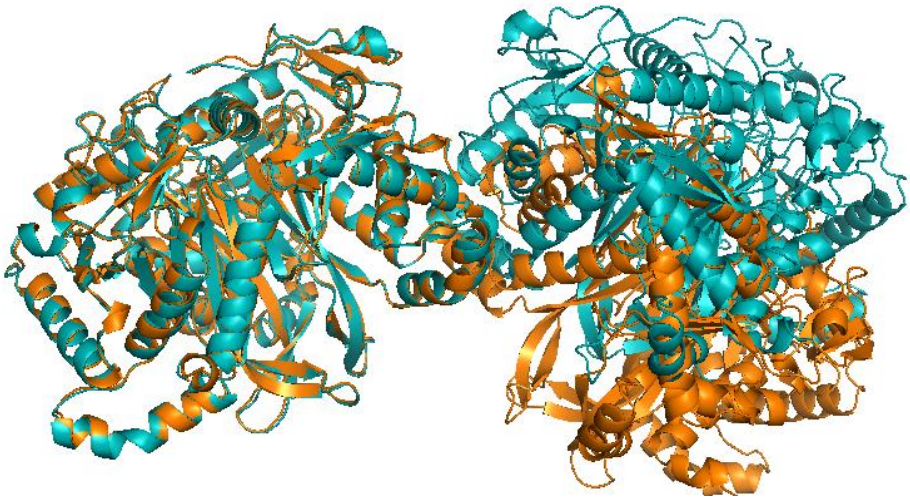
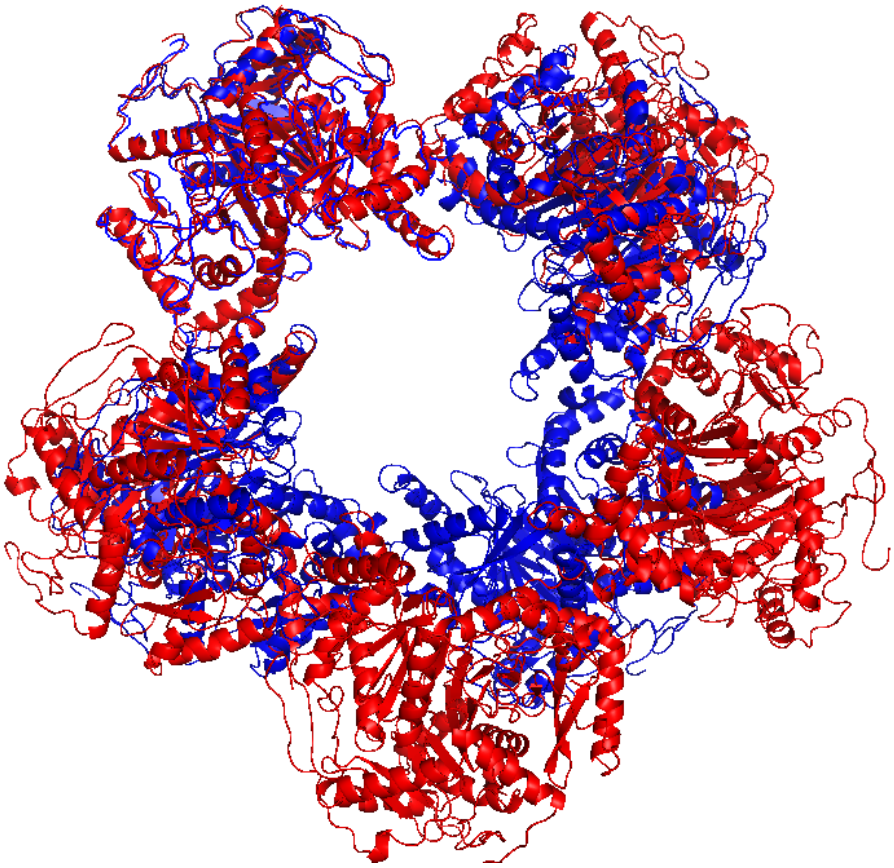


**Figure 5.16 Overlaid structures of 2 adjacent trimers from a cubic 24-mer and a dodecahedral 60-mer**. The trimers on the left have been superimposed over each other, the *A. vinelandii* ribbon diagram is shown in orange and the *G. stearothermophilus* in blue.

A)



B)

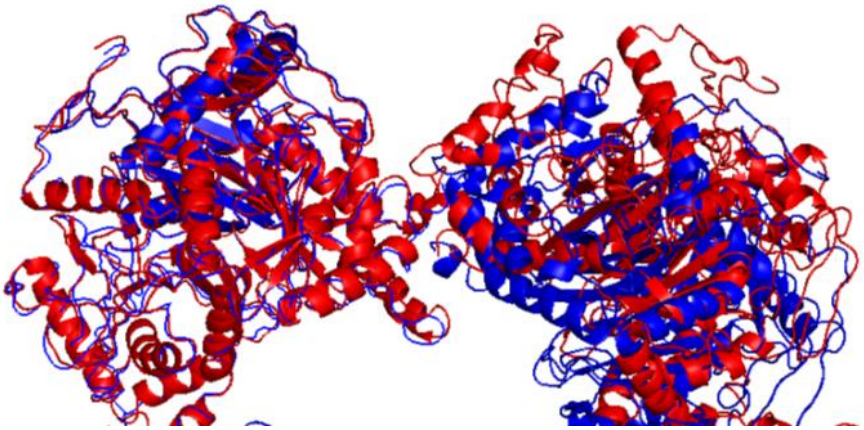


**Figure 5.17 The square and pentagonal faces of *the Tp. acidophilum* 42-mer.** A) The top left trimer of the square ribbon diagram (blue) and pentagonal (red) faces have been aligned with one another. B) Illustrates the different angles at which the trimers interact with each other viewed from the centre of the face with the aligned trimer of the left.

The *Tp. acidophilum* E2 trimers located on the square and pentagonal faces of the 42-mer interact at differing angles (Figure 5.17). Figure 5.18 illustrates how the differing angles of the two trimers relative to each other may influence the number of possible inter-trimer interactions. The trimers located on a pentagonal faces are arranged so that there is a possibility of an ionic bond forming between the arginine of one trimer and an aspartic acid on the second (approximately 4.5Å apart), and *vice versa* (Figure 5.18A). The same interaction appears unlikely between two adjacent trimers located on a square face, as these two residues are now located approximately 10Å apart.

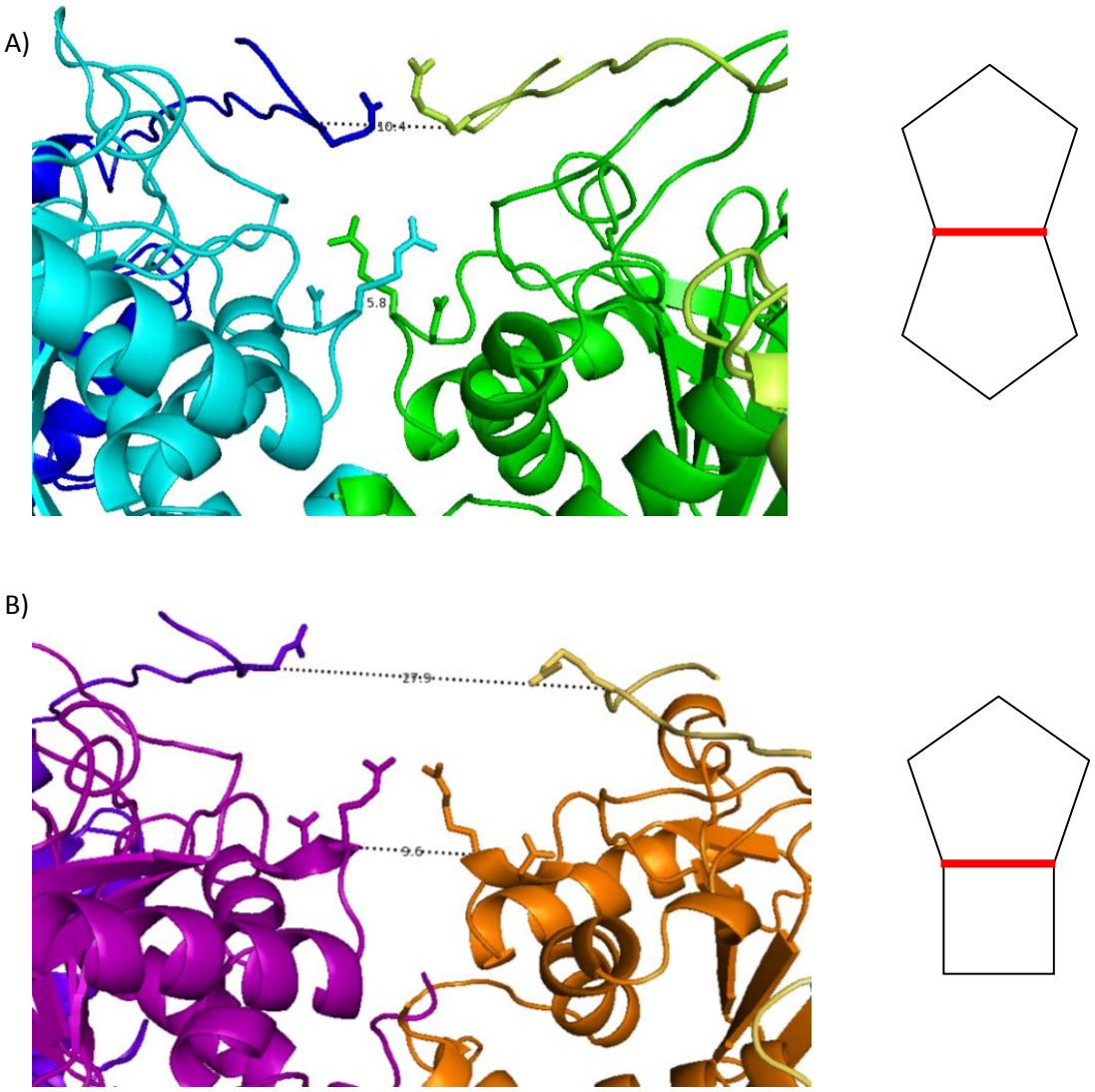



**Figure 5.18 Trimer-trimer interfaces on A) a pentagonal and B) a square face of the *Tp. acidophilum* E2 42-mer**. The pentagonal-face trimers are shown in shades of blue and green and the square-face trimers in shades of purple and orange. The side chains of key residues have been included on these images but it should be noted that the exact location of these residues is unclear. Measurements are shown between the polypeptide backbone bones. The line diagram next to each image indicates the location of the two trimers on either the square or pentagonal faces.

The 4Å resolution of this structure and the 60-mer prevents the exact identification of the residues that are responsible for determining the angle at which the trimers interact. Interactions between the anchor residue and the hydrophobic pocket, or surface regions at the inter-trimer interface may influence the angle at which the trimers interact with one another, and ultimately govern the overall structure of the E2. A higher resolution structure of the 42-mer and 60-mer might confirm which residues and interactions are important; however, this may not be feasible, so a second option is to mutate the amino acids predicted to contribute side chains to the inter-trimer interactions and determine the effect on the overall structure.

The E2 trimers within the 42-mer are all identical, but the individual monomers occupy differing environments depending upon their location within a square or pentagonal face; this variation may go beyond the quasi-equivalence principles previously suggested, by Izard et al. (1999), to govern the assembly of the OADHC E2 proteins. Quasi-equivalence suggests that all protein subunits are linked by the same type and number of bonds, but these bonds may be deformed slightly (Caspar and Klug, 1962). The interactions between the trimer of the 24-mer and 60-mer structures are described as quasi-equivalent; however the interaction between the non-Euclidean 42-mer trimers cannot be described as quasi-equivalent. The bonds between the 42-mer trimers are not merely distorted, in some cases they are completely different, some interactions are only possible between monomers located within pentagonal faces and not between monomers located on a square face.

The *Tp. acidophilum* 42-mer structure cannot be predicted based on quasi-equivalence and the rules of Euclidean symmetry; but it can, however, be explained by another argument that subunits are not obliged to occupy identical environments and form identical interactions, rather they assemble into the lowest energy structure possible (Caspar and Klug, 1962, Caspar, 1980).

***5.3.4.6 Thermostability of the E2cat protein***

Evidence suggests that protein thermostability is not achieved by one method, but rather through a number of small changes. A range of structural features have been proposed to contribute to protein thermostability, although their contribution and usage varies depending on the organism and protein, these adaptations include:

- Decreased conformational flexibility
- Decreasing fraying by reduced terminal extensions and additional terminal interactions
- Increased packing efficiency
- Increase in salt bridge networks
- Increased hydrogen bonding
- Extra disulphide bonds
- Amino acid substitutions
    - More hydrophobic and helix stabilising residues
    - Fewer exposed thermolabile amino acids

(Radestock and Gohlke, 2010, Daniel et al., 2008, Li et al., 2005)

The E2cat is a hyperthermostable protein but the structural basis of this thermostability is not clear from this 4Å model. Reducing the flexibility of the full length E2 is not a feasible option for this protein as the flexible linker regions are key to enzyme functionality; instead one possibility is that the E2 catalytic domain has evolved to become hyper-thermostable and therefore capable of resisting the denaturing forces exerted on it (Chapter 4.4). It is not currently possible to prove this hypothesis as thermostability data from mesophilic and thermophilic E2cat and full length E2 protein are not available.

Despite the variety of potential stabilising interactions referred to above, one prevalent feature of thermophilic and hyper-thermophilic proteins is that they tend to have more charged residues, resulting in the possibility of forming extra ionic bonds. These interactions increase in strength at higher temperatures as the energy required to desolvate a charge residue is lower, due to increased water entropy, but the Coulombic interaction between opposite charges within an ionic bond remains the same; therefore, the formation of ionic interactions becomes increasingly favourable as temperatures increase. Hyper-thermophilic enzymes also utilise a network of ionic bonds. The formation of an ionic bond requires the desolvation of two charged residues in order to

create one bond, but the addition to the interaction of a third charge to form a network requires the desolvation of only one charged residue; given that a full Coulombic interaction is gained, ionic networks are highly favourable with respect to thermostability.

The number of ionic interaction in the *Tp. acidophilum* E2cat cannot be accurately determined, and a higher resolution structure would be required to determine the positions of the side chains and thus to identify ionic interactions. The percentage of charged residues in full-length E2 sequences from mesophilic through to hyper-thermophilic organisms can, however, be compared; there is a positive correlation between the percentage of charged residues and the optimal growth temperature of the source organism (Figure 5.19). The *Tp. acidophilum* E2 protein contains a higher percentage of charged residues than the mesophilic *E. coli* PHD E2 but less than the hyper-thermophilic E2 from *Pyrobaculum aerophilum*.

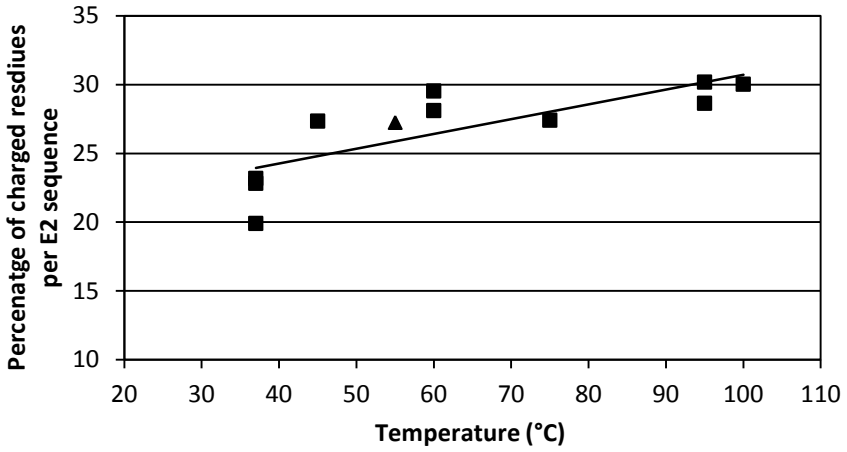
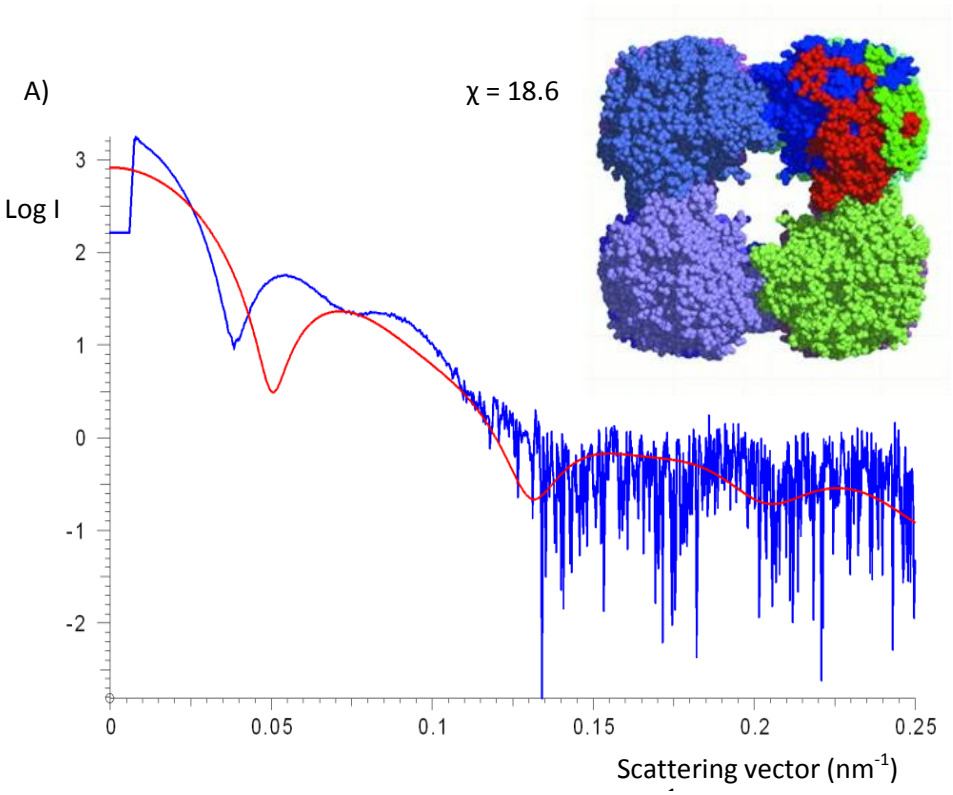


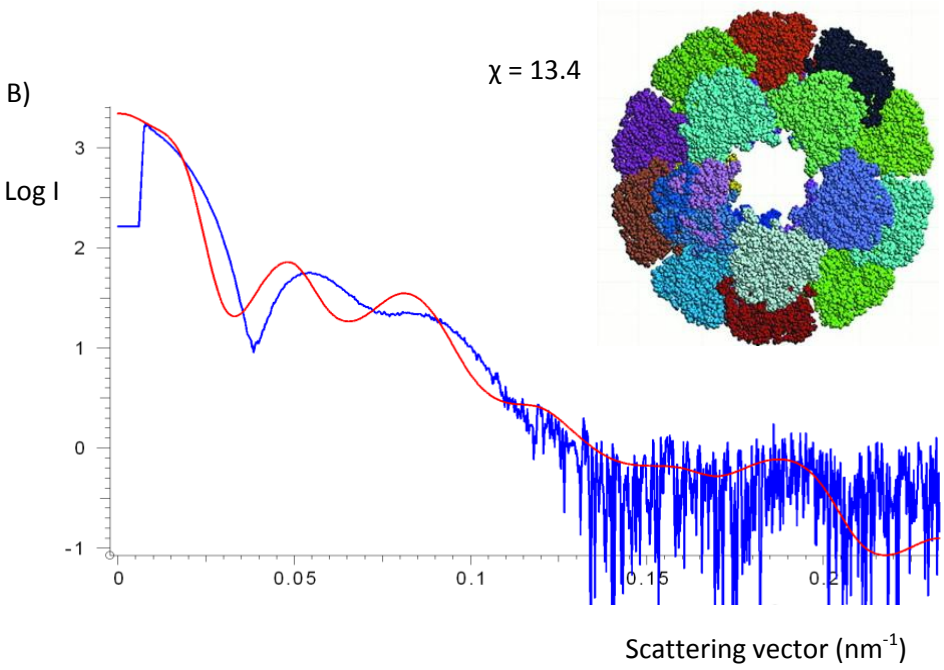

**Figure 5.19 The percentage of charged residues per full-length E2 as a function of the growth temperature of the source organism.** The total number of positive and negative residues was calculated as a percentage of the total number of resides in each E2. The temperature corresponds to the optimal growth temperature of the organism. The E2 proteins sequences are taken for the following mesophilic to hyper-thermophilic organisms: human, cow, *E. coli, Ferroplasma acidarmanus, G. stearothermophilus, Tp. acidophilum* (shown as a triangle)*, Picrophilus torridus, Sulfolobus islandicus, Aeropyrum pernix, Pyrobaculum calidifontis*, and *Pyrobaculum aerophilum.*

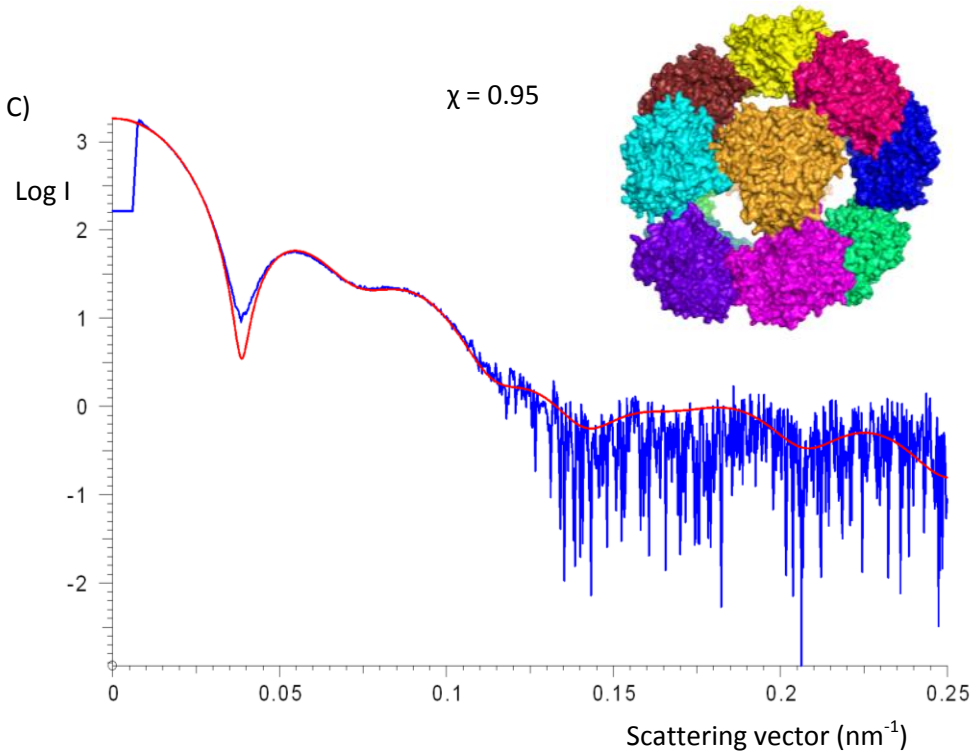
### 5.3.5 Small-angle X-ray scattering

SAXS was used to determine the E2cat structure in solution, and hopefully to verify that the novel 42-mer X-ray crystallographic structure was not an artefact of crystallisation. The SAXS data indicated that the E2cat protein, in 50mM Tris-HCl pH 8.8 and 100mM NaCl, was monodisperse, with pronounced subsidiary maxima, characteristic of a hollow globular protein structure. The Guinier plot determined the radius of gyration (*R*g) to be 85 ±3 Å, and the P(*r*) (distance distribution) plot determined the maximum linear dimension ($D_{max}$) value to be 220 ± 10 Å; these values imply that the E2cat is a distorted sphere structure, as the $D_{max}$ value is not double the *R*g value, as would be expected for a perfect sphere. The molecular weight determined from these experiments was 950 ± 90 kDa, equivalent to 37.3 ± 3.5 monomers.

Figure 5.20 shows the experimental scattering pattern compared to the predicted scattering pattern produced by the 24, 42 and 60-mer structures (PDB codes: 1EAF, 3RQC, and 1B5S respectively); only the predicted 42-mer pattern fitted the experimental data ($\chi$ = 0.95). A hollow spherical *ab initio* bead model was produced from the SAXS data that superimposed well with the X-ray crystal structure ($\chi$ = 1.6) (Figure 5.20D); this indicates that the 42-mer crystal structure is not an artefact of crystallisation.


A)
χ = 18.6
Log I
3
2
1
0
-1
-2
0
0.05
0.1
0.15
0.2
0.25
Scattering vector (nm-1)


B)
χ = 13.4
Log I
3
2
1
0
-1
0
0.05
0.1
0.15
0.2
Scattering vector (nm-1)



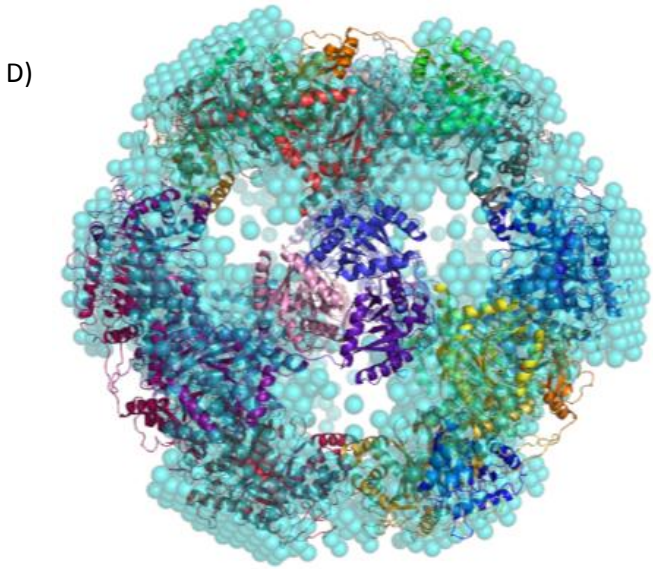




**Figure 5.20 Small-angle X-ray scattering data and analysis**. A-C) illustrates the X-ray scattering data in blue along with predicted scattering pattern in red for a 24-mer, 60-mer and 42-mer structure respectively; the chi values for fit of the model to the experimental data are shown above the graph. Image D shows the low resolution 42-mer *ab initio* bead model (cyan) superimposed over the crystal structure.

### 5.3.6 Sedimentation equilibrium analytical ultracentrifugation

The sedimentation equilibrium analytical ultracentrifugation experiments were performed and analysed by Dr. Jacqui Marshall (University of Bristol). The molecular weights of the purified E2 and E2cat proteins (Figure 5.21) in 50mM Tris-HCl, pH 8.8, and 100mM NaCl, were determined by fitting the equilibrium data from two speeds and three protein concentrations (when possible) to a single ideal species model (example shown in Figure 5.22).

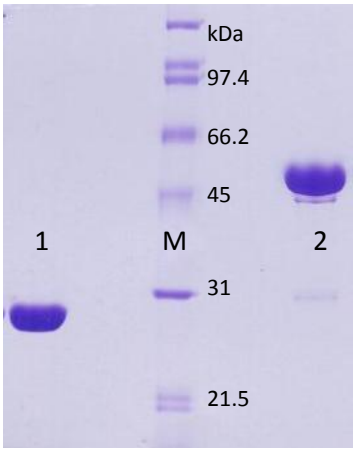



**Figure 5.21 The E2 and E2cat protein samples used for centrifugation.** E2cat protein is shown in lane 1 and the full-length E2 protein in lane 2. The standard protein markers, with their molecular weights (kDa) are shown in lane M.

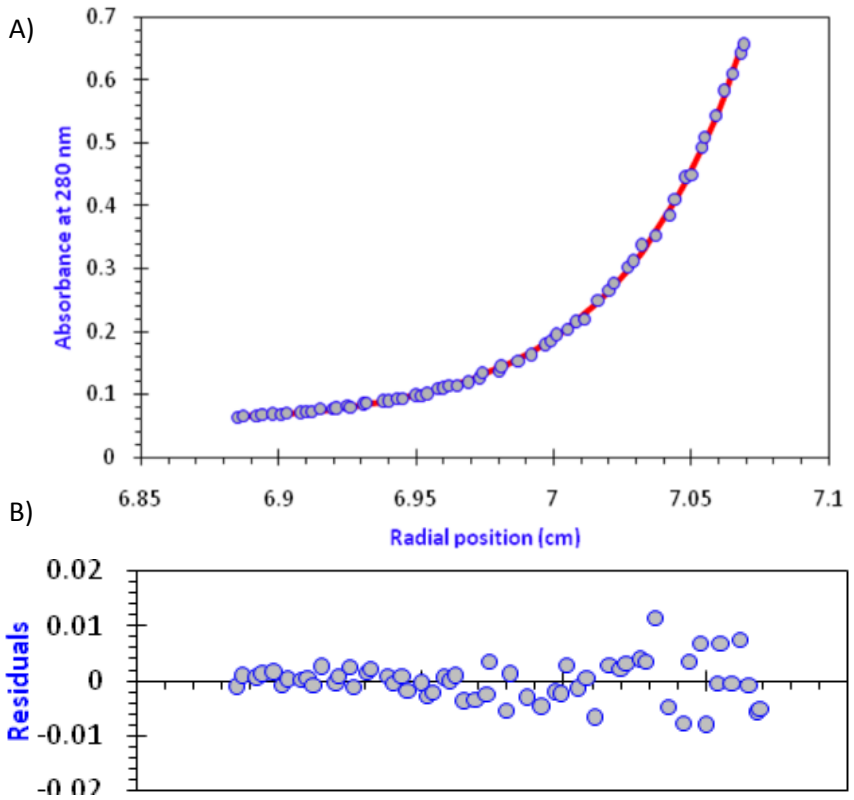



**Figure 5.22 Sedimentation equilibrium analytical ultracentrifugation of the E2cat protein**. A) The blue dots are the recorded $A_{280}$ values with radial distance across the cell after 48 h at 5000 r.p.m. The red line is the fitted model for a single ideal species with the residuals from fitting the data to this model shown in B.

| **Protein** | **Expected $M_w$ 42-mer (MDa)** | **Calculated $M_w$ (MDa)** | |
|---|---|---|---|
| | | **3k r.p.m.** | **5k r.p.m.** |
| **E2cat** | 1.07 | 1.20, 1.18, 1.25<br>**Average = 1.21** | 1.02, 1.07, 1.08<br>**Average = 1.08** |
| **Full-length E2** | 1.95 | 1.93, 1.89, 2.10<br>**Average = 1.94** | 1.58, 1.52, 1.74<br>Not random residuals |

**Table 5.4 The molecular weights of the E2 and E2cat protein determined by sedimentation equilibrium ultracentrifugation.** The molecular weights of the E2 and E2cat proteins are shown for 3000 and 5000 r.p.m, at three concentrations listed from highest to lowest along with the average $M_w$ for each speed. The E2 5000 r.p.m result does not have an average due to the non-random residual produced during the fitting of the data to the single ideal species model.

The E2cat 3k r.p.m and 5k r.p.m results and the full-length E2 3k r.p.m result fit the single ideal species model; therefore, an accurate molecular weight can be determined. The E2cat 3k r.p.m and 5k r.p.m results were combined resulting in a molecular weight of 1.08 MDa ±0.02, equivalent to 42.2 (±0.8) subunits. The 3k r.p.m results was used to determine the molecular weight of the E2 protein to be 1.94 (±0.05) MDa, this equates to 41.8 (±1.1) subunits (Table 5.4); the 5k r.p.m results for the E2 have not been included as they do not fit the model; this may be due to the steepness of the curve at this speed making it difficult to fit the model accurately to the data points.

Sedimentation equilibrium analytical ultracentrifugation is a method that determines the molecular weight of a protein independently of its structure; this result therefore strongly supports the previous conclusion that the *Tp. acidophilum* E2cat protein assembles into a novel structure composed of 42 monomers. This method also confirms that the removal of the flexible linker region does not affect how the E2 trimers interact, as both the full-length E2 and the E2 catalytic both assemble into 42-mers.

## 5.4 Conclusion

In this chapter, several complementary techniques have been used to determine the structure of the *Tp. acidophilum* E2cat protein. X-ray crystallography, small-angle X-ray scattering, and sedimentation equilibrium analytical ultracentrifugation all confirm that the *Tp. acidophilum* E2cat protein assembles into a 42-mer structure; analytical ultracentrifugation also proves that the full-length E2 protein assembles into a single structure with a molecular weight consistent with that of a 42-mer. All currently known E2 structures are Euclidean shapes, either cubic 24-mers with square faces or dodecahedral 60-mers with pentagonal faces. The *Tp. acidophilum* E2 structure is an intermediate between the 24-and 60-mer structures, comprising 42 monomers with both square and pentagonal faces; despite this apparent intermediate structure only a 42-mer is formed, rather than any other shape. The data presented in this chapter contradict the previous sedimentation-velocity analytical ultracentrifugation experiments conducted by Heath et al. (2007) that determined that the E2 was a 24-mer. The discrepancy between the two results may be due to proteolysis of the E2; if the flexible linker region between the PSBD and the catalytic domain were cleaved, the resulting protein would have a monomer $M_w$ = 26.8 kDa, and a 42-mer $M_w$ = 1.13 MDa, this is close to the 1.11 MDa expected for an intact E2 24-mer. The E2 protein samples were not analysed by SDS-PAGE after the experiments of Heath et al. (2007), so it is not possible to confirm this theory.

The structural adaptations that cause the *Tp. acidophilum* E2 to form a 42-mer rather than 24-or 60-mer are not understood; currently there are not enough structures available from differing organisms, temperatures and assemblies, to be able to predict which changes are neutral or influence thermostability and/or assembly.

# Chapter 6: The effects of mutating and removing the putative E2 anchor residue on the assembly of the 42-mer

## 6.1 Introduction

The amino acid side-chains involved in the *Thermoplasma acidophilum* E2 catalytic domain trimer-trimer interactions are of interest, due to the hyper-thermostability of this protein and its ability to form a novel 42-mer. The C-terminus of the E2 catalytic domain has been shown to be important in the formation of trimer-trimer interactions; the addition of a C-terminal his-tag to the *E. coli* succinyltransferase disrupted inter-trimer interactions, resulting in E2 trimers that no longer associated with one another (Knapp et al., 2000). One residue in particular, an "anchor residue", is thought not only to be important in the formation of trimer-trimer interactions, but also affect the angle at which the trimers interact, and this in turn influences the multimeric structure of the E2 assembly (Izard et al., 1999). The E2 anchor residue is a hydrophobic amino acid, usually a valine or leucine, located a few amino acids from the E2cat C-terminus; the hydrophobic anchor residue side-chain extends from one trimer into a hydrophobic pocket on the adjacent trimer, and *vice versa,* tethering neighbouring trimers to each other. Unusually, the *Tp. acidophilum* E2cat appears to have an isoleucine as its anchor residue; the aim of this chapter was to determine how either mutating or removing this putative anchor residue affected the 42-mer assembly.

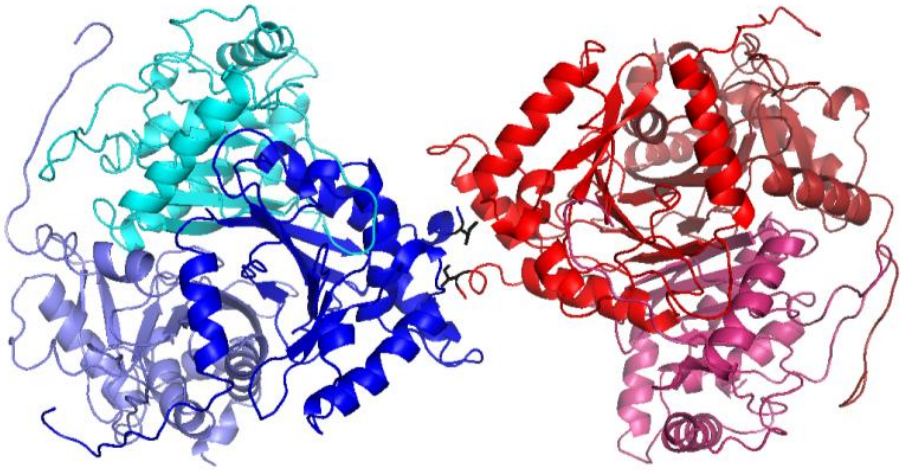


**Figure 6.1 A ribbon diagram displaying the location of the putative anchor residue (E2-Ile 397) at the interface between two *Tp. acidophilum* E2cat trimers.**

## 6.2 Methods

### 6.2.1 Identification of a putative anchor residue

The *Tp. acidophilum* E2cat sequence was aligned and the structure compared with the following 24-mer and 60-mer E2s: *G. stearothermophilus* PDHC, *E. coli* OGDHC, *A. vinelandii* PDHC, and bovine branched-chain OADHC. Two possible Ile anchor residues were identified from the E2 sequence (whole E2: Ile396 and 397; E2cat: Ile 220 and 221,); the two residues were mapped onto the *Tp. acidophilum* E2cat structure to determine which was the more likely candidate. The sequence below shows the location of the two possible Ile anchor residues.

***Tp. acidophilum* whole E2**

MYEFKLPDIGEGVTEGEIVRWDVKEGDMVEKDQDLVEVMTDKVTVKIPSPVRGKIVKILYREGQVVP
VGSTLLQIDTGEEAPVQQPAGRAESTVQVAEVKQVPLPEVSGHVLASPAVRRIARENGIDLSKVGGTG
EGGRVTLDDLERYMKSPAPSPAPSAGKAEAVHTAPQIPAQKPAPGREEILEMHGLRRIIFDKMTKAKQ
IMPHFTVMEEVDVTSMVSILDSAKARNRKVTVTGFLARIVPSILKQYPYLNAIYDETRRVYILKKYYNIGI
AVDTPDGLNVFVIKDADRKSMVEISAEISDKASRARENKLQLDEVQDSTFTITNVGTIGGIMSTPIINYP
EVAILGVHRILEREGRKYMYLSLSCDHRLIDGAVATRFIVDLKKVIEDPNA**II**YEI

### 6.2.2 Site-directed mutagenesis and truncation of the E2 and E2 catalytic domain

Site-directed mutagenesis was used to mutate the putative E2cat anchor residue (E2cat-Ile 221) to either a valine or leucine, following the method described in Chapter 2.2.1 (primers shown in Table 6.1). In addition, the C-terminal anchor residue was removed from both the E2 and E2cat protein by the insertion of a premature stop codon before the final 5 amino acids (IIYEI); the stop codon (TGA) was inserted by site-directed mutagenesis using the primers shown in Table 6.1. The presence of the correct mutations was determined by sequencing the E2 gene in pET28a and E2cat in pET24a, using the vector specific T7F and T7R primers by Source BioScience Plc (Oxford, U.K.).

| | | |
|---|---|---|
| **Original** | | 5′GAGGATCCGAACGCTATA***ATC***TACGAGATC**TGA**CTC3′ |
| **Valine** | **Forward** | 5′GAGGATCCGAACGCTATA**GTC**TACGAGATC**TGA**CTC3′ |
| | **Reverse** | 5′GAG**TCA**GATCTCGTA**GAC**TATAGCGTTCGGATCCTC3′ |
| **Leucine** | **Forward** | 5′GAGGATCCGAACGCTATA**CTC**TACGAGATC**TGA**CTC3′ |
| | **Reverse** | 5′GAG**TCA**GATCTCGTA**GAG**TATAGCGTTCGGATCCTC3′ |
| **Truncation** | **Forward** | 5′GAGGATCCGAACGCT**TGA**ATAATCTACGAGATC**TGA**C3′ |
| | **Reverse** | 5′G**TCA**GATCTCGTAGATTAT**TCA**AGCGTTCGGATCCTC3′ |

**Table 6.1 E2 and E2cat mutagenesis primers**. Primers designed to mutate the E2/E2cat anchor residue isoleucine 397/221, or to truncate the sequence before isoleucine 396/220. Stop codon sequence in red, Ile anchor residue in green, mutation in blue and early stop codon in purple.

### 6.2.3 Expression of the mutated and truncated E2 and E2 catalytic domain proteins

The truncated E2 and E2cat proteins were expressed following the same method as the non-truncated E2 and E2cat proteins, described in Chapters 2.3.1 and 3.2.2, respectively. A range of different expression conditions were tried in order to express the conservatively mutated E2cat protein in a soluble form, including using Overnight Express™ Instant TB medium (Novagen, Merck, U.K.), and altering the temperature (between 20-37°C) and length of growth (between 4-20h) after induction in LB medium.

### 6.2.4 Purification of the truncated E2 and E2 catalytic domain proteins

The truncated E2 (E2-trunc) was purified by heat precipitation followed by nickel affinity chromatography (Chapter 2.3.3). The truncated E2cat (E2cat-trunc) was purified by the 4 step purification method established for the E2cat protein in Chapter 4.2.2: 10 min heat precipitation at 55°C, DNA precipitation by titration of the cell lysate with PEI, followed by anion exchange and gel filtration chromatography. The E2-trunc was also purified by this method, instead of nickel affinity chromatography, when higher purity was required for structural studies.

### 6.2.5 Determining the size and multimeric nature of the E2-trunc and E2cat-trunc proteins

#### *6.2.5.1 Purification and size determination by gel filtration*

The final purification step and size determination experiments were carried out as described in Chapter 2.3.3.3 using a Superdex 200, 10-300GL column equilibrated with 50mM Tris-HCl, pH 8.8, and 100mM NaCl.

#### *6.2.5.2 Sedimentation equilibrium analytical ultracentrifugation analysis of the E2-trunc and E2cat-trunc proteins*

All experiments were performed and results analysed by Dr. Jacqui Marshall (University of Bristol) using an An60Ti rotor in a Beckman XL-A ultracentrifuge at 20°C, as described in Chapter 5.2.5. The experiments were performed using the same buffer as the previous E2 and E2cat experiments: 50mM Tris-HCl, pH 8.8, and 100mM NaCl.

### 6.2.6 Determination of the structure of the truncated E2 catalytic domain

#### *6.2.6.1 Small-angle X-ray scattering*

The SAXS data were collected at the X33 beam-line, EMBL Hamburg Outstation, (DORIS-III storage ring at Deutsches Elektronen-Synchrotron (Roessle et al., 2007)) using an automatic sample changer (Round et al., 2008); the SAXS data were processed and analysed by Dr. Dmitri Svergun, Dr. Peter Konarev and Dr. Gundolf Schenk (EMBL, Hamburg, Germany) and Dr Jean van den Elsen (University of Bath, U.K.). The results were analysed by the method described in Chapter 5.2.4.

#### *6.2.6.2 E2cat-trunc X-ray crystallography*

An Art Robbins Nano dispenser robot was used to set up the following 96 well condition screens: PGA-LM, Crystal Screen 1 and 2, Morpheus, PACT and JCSG (Molecular dimensions, Suffolk, U.K.); each of the 96 conditions had 3 sitting drops with a differing ratio of protein to well solution: 1:2, 1:1 and 2:1 protein (~10mg/ml) to well solution, totalling 300nl. The plates were sealed with a plastic sheet and stored at 16°C. The screens were checked every two days for two weeks, and then weekly for several months.

The crystals were optimised following the method detailed in Chapter 5.2.3.2. A crystal grown in 8% (w/v) polyethylene glycol (PEG) 4k (Fluka, Sigma-Aldrich), 3% (w/v) Poly-glutamic acid (PGA-LM) (Molecular dimensions, U.K.), 0.1M Na cacodylate and pH 6.5 was soaked in the following cryoprotectant before being flash frozen: 8% (w/v) PEG 4k, 3% (w/v) PGA, 0.1M Na cacodylate, pH 6.5 and 30% glycerol. The resultant crystal was used to collect X-ray diffraction data at the Diamond Light Source (Oxfordshire, U.K.) on an ADSC Q315 3x3 CCD detector on station I04 (at wavelength $\lambda$=0.9702 Å). The data were collected and processed using the HKL2000 package by Dr Susan Crennell (University of Bath).

The E2cat-trunc model was built and refined by Dr. Jean van den Elsen (University of Bath). The E2cat and E2cat-trunc monomer crystal structures were compared using the DaliLite programme (Holm and Park, 2000).

## 6.3 Results and discussion

### 6.3.1 Mutation of the E2cat anchor residue

The putative E2cat anchor residue (E2cat-Ile 221) was mutated into a valine and a leucine by quick-change site-directed mutagenesis. Recombinant expression of the E2cat (Val 221) and E2cat (Leu 221) resulted in insoluble protein in all conditions tried (Figure 6.2).

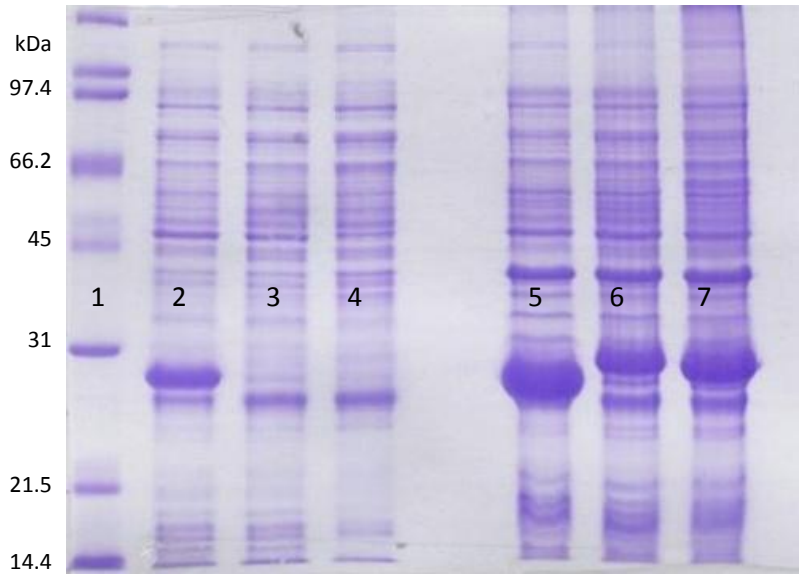



**Figure 6.2 SDS-PAGE of the soluble and insoluble cell lysate expressing the wild-type and mutated E2cat proteins**. Lane 1: standard protein markers, with their molecular weights (kDa). Lanes 2 and 5 contain the soluble and insoluble non-mutated E2cat cell extract proteins. Lanes 3 and 6, and 4 and 7, contain the soluble and insoluble E2cat (Ile 221 Val) and E2cat (Ile 221 Leu) cell lysate respectively.

### 6.3.2 Removal of the anchor residue by truncating the E2 and E2cat proteins

The putative anchor residue (E2-Ile 397/ E2cat-Ile221) was removed by the insertion of a stop codon (TGA), before the final five C-terminal amino acids (IIYEI); the insertion of the stop codon was confirmed by sequencing. The sequence of the E2 protein's N-terminal lipoyl domain and PSBD are unaffected by this truncation.

### 6.3.3 Expression and purification of the truncated E2 and E2cat proteins

#### *6.3.3.1 Expression of the truncated E2 and E2cat proteins*

The E2-trunc was expressed solubly with an N-terminal his-tag from pET28a. TheE2cat-trunc protein was expressed in a partially soluble form without a his-tag from pET24a when the cells were induced with IPTG (Figure 6.3).

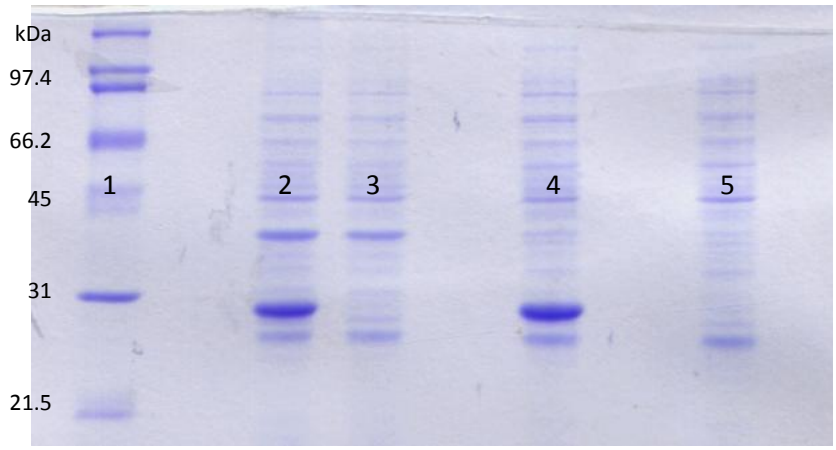



**Figure 6.3 SDS-PAGE comparing the soluble and insoluble cell lysate proteins when the *E. coli* cells containing the E2cat-trunc-pET24a vector were induced and not induced with IPTG.** Lane 1: standard protein markers, with their molecular weight values (kDa); lanes 2 and 3: Insoluble protein from induced and non-induced cell cultures respectively. Lanes 4 and 5: soluble protein from induced and non induced cell cultures.

### *6.3.3.2 Nickel affinity chromatography purification of truncated E2*

The E2-trunc protein was purified to between 85 and 95% purity, by heat-precipitation, and nickel-affinity chromatography (Figure 6.4); although some of the E2-trunc protein eluted during the low concentration imidazole wash steps, the majority eluted in the first two His-elute buffer 1ml washes.

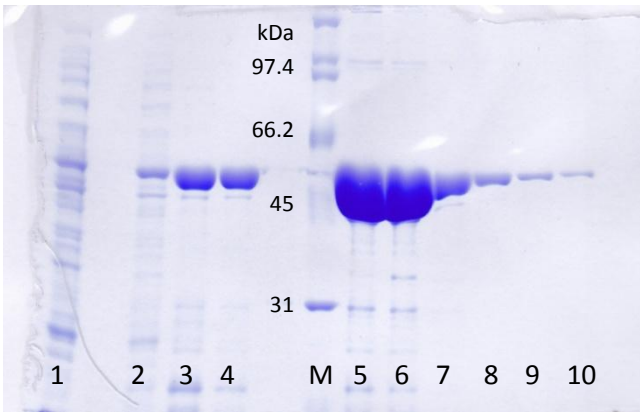



**Figure 6.4 SDS-PAGE of E2-trunc purification by nickel affinity chromatography.**
Lane 1: flow through protein that did not bind to the column; lanes 2: the his-bind buffer wash, lanes 3 and 4: the 5% and 10% (v/v) His-elute fractions, respectively; and lanes 5 to 9: his elute wash collected in 5 separate 1ml fractions. The standard protein markers are in lane M and are labelled with their molecular weight values.

### *6.3.3.3 Anion exchange chromatography purification of the E2cat-trunc*

An example E2cat-trunc anion exchange chromatography trace is shown in Figure 6.5. Cell lysate was loaded onto the column in 50mM Tris-HCl, pH 8.8, in this buffer the E2cat-trunc protein did not bind to the anion exchange column, but the majority of contaminating *E. coli* proteins did. The flow through fractions (A3- A6 in Figure 6.6) were pooled and concentrated before further purification by gel filtration chromatography.

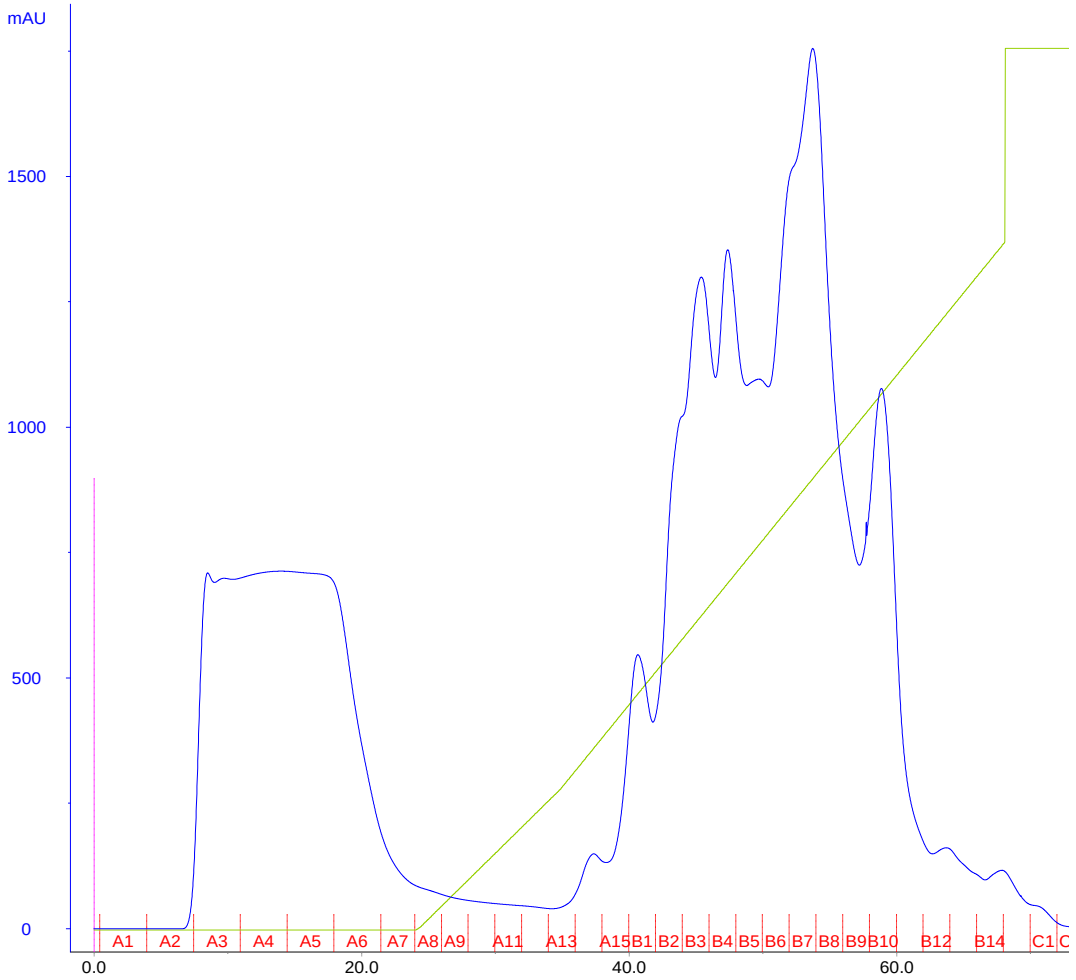



**Figure 6.5 Anion exchange trace of the E2cat-trunc protein purifcation.** The $A_{280}$ readings are shown in blue, the fraction numbers are shown along the x-axis, and the concentration of NaCl is shown in green, increasing from 0 to 1M.

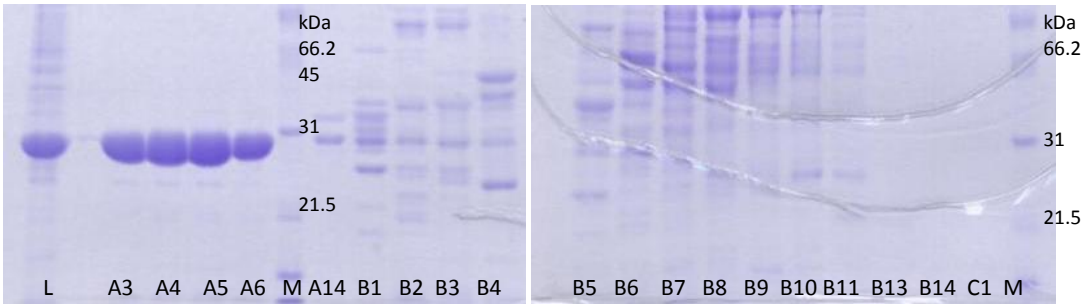



**Figure 6.6 SDS-PAGE of anion exchange fractions**. The lanes are labelled by the fraction number corresponding to the fractions shown in Figure 6.4; the standard protein markers, with their molecular weight values (kDa) are shown in the lanes labelled M and the lane labelled L is the load fraction.

### *6.3.3.4 Gel filtration chromatography of the truncated E2 and E2cat*

Size exclusion gel filtration chromatography was used as a final purification step, and also used to determine the molecular weight of the E2-trunc and E2cat-trunc proteins (both in 50mM Tris, pH 8.8, and 100mM NaCl).

The E2-trunc protein eluted from the column in a single peak at 11.23 ml (Figure 6.7), this elution volume equates to an approximate molecular weight of 340 kDa, the equivalent to 7 E2-trunc subunits.

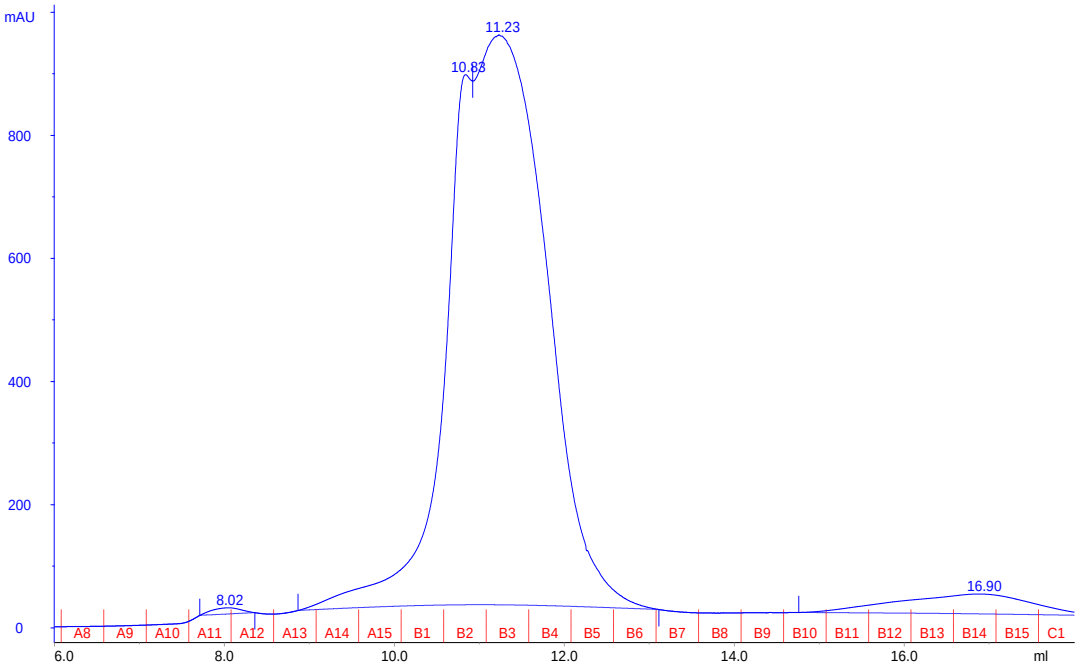



**Figure 6.7 E2-trunc Gel filtration trace**. The $A_{280}$ readings are shown in blue, and the 0.5ml fraction shown in red along the x-axis. No protein peaks were observed prior to those shown.

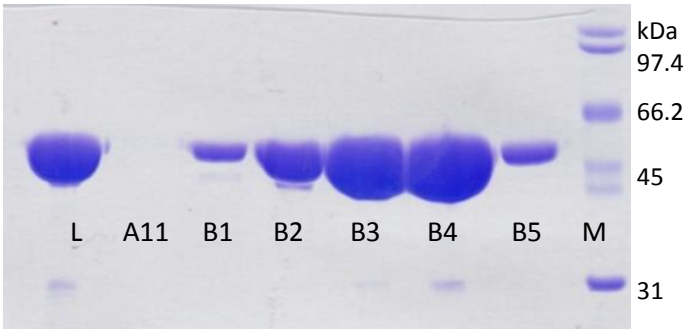



**Figure 6.8 SDS-PAGE of the E2-trunc gel filtrations fractions.** The lanes are labelled with the fraction number corresponding to Figure 6.7 The lane labelled L contains the load sample and the lane labelled M contains broad range molecular weight markers along with their molecular weights (kDa).

The E2cat-trunc protein eluted from the gel filtration column at 15.25ml (Figure 6.9); the molecular weight of the protein in this peak was determined to be approximately 61 kDa, the equivalent to 2.45 E2cat-trunc monomers.

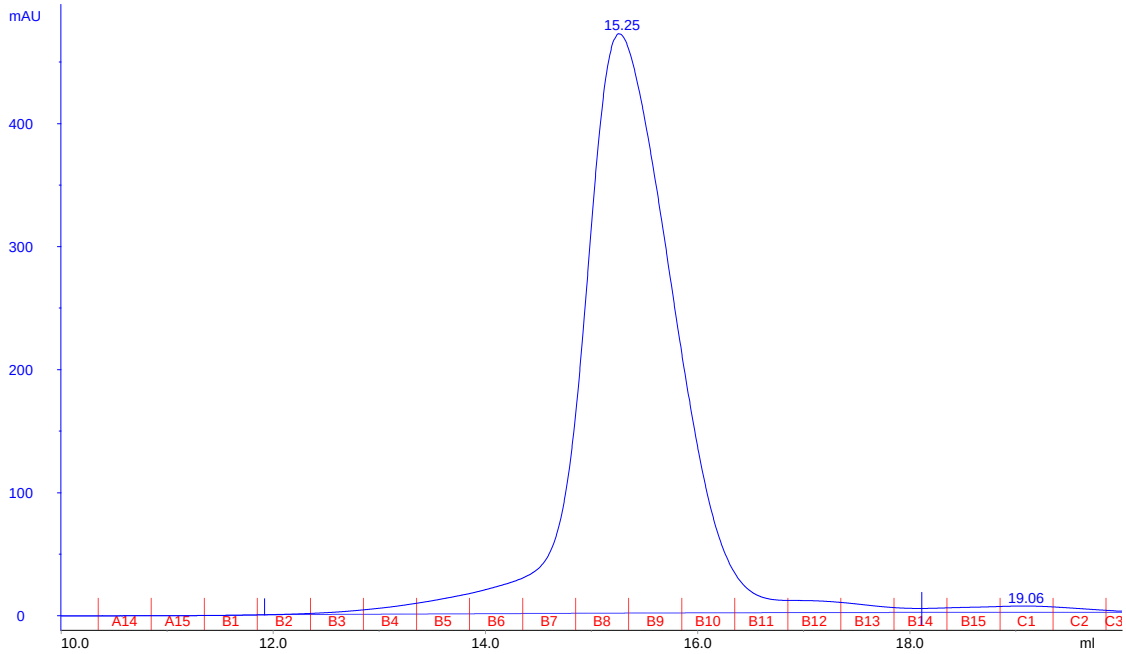



**Figure 6.9 E2cat-trunc gel filtration trace.** The $A_{280}$ readings are shown in blue, and the fraction numbers are shown along the x-axis.


**Figure 6.10 SDS-PAGE of E2cat-trunc gel filtration fractions.** The lanes are labelled by fraction number corresponding to the fractions shown in Figure 6.9. The lane labelled M contains the standard protein markers, with their molecular weight values (kDa) and the lane labelled L contains the proteins loaded onto the column

The E2-trunc and E2cat-trunc proteins did not elute from the column in the void volume like the non-truncated versions of these proteins. The E2-trunc gel filtration results imply that this protein assembles into a structure comprising 7 monomers; however, this is

unlikely to be an accurate molecular weight, as the flexible linker regions of the E2 protein extend out from the catalytic domain, creating a larger hydrodynamic radius than would be expected for a protein of equal molecular weight. The E2cat-trunc data suggests that this protein is forming either dimers or trimers. The conclusion drawn from the gel filtration data is that the truncation of the C-terminal $3_{10}$ helix results in E2 and E2cat proteins that no longer form 42-mers.

### 6.3.4 Sedimentation equilibrium analytical ultracentrifugation of the truncated E2 and E2cat proteins

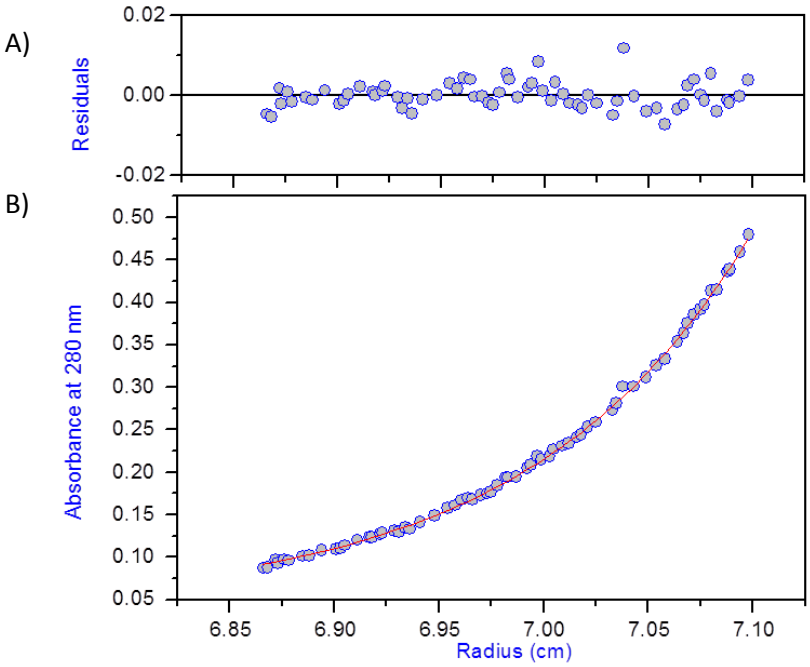
The E2-trunc and E2cat-trunc proteins do not appear to form 42-mers. The molecular weights of these proteins determined by gel filtration are unlikely to be accurate, as this technique can be influenced by the shape of a protein. Therefore, equilibrium analytical ultracentrifugation (AUC) was used to accurately determine the molecular weights of the two proteins and thus elucidate the multimeric nature. These experiments were performed on E2-trunc and E2cat-trunc protein in 50mM Tris-HCl, pH 8.8, and 100mM NaCl, the same buffer as was used for the non-truncated E2 and E2cat protein AUC experiments.

#### *6.3.4.1 Determination of the molecular weight of the truncated E2 protein*

The molecular weight of the E2-trunc was determined at two speeds (7.5k and 9.5k r.p.m.) and at 3 protein concentrations (1.3, 0.8, 0.4mg/ml) (Table 6.2). The samples were run on an SDS-PAGE after centrifugation to ensure proteolysis of the flexible linker regions had not occurred during the experiment (Figure 6.12).

| | **$M_w$ (kDa; data fitted to a single ideal species model)** | | | |
|---|---|---|---|---|
| | **Channel A** | **Channel B** | **Channel C** | **Global fit** |
| **Protein concentration (mg/ml)** | 1.3 | 0.8 | 0.4 | |
| **7k r.p.m. 24 h** | 118.9 | 129.1 | 135.7 | 123.5 |
| **9.5k r.p.m. 30 h** | 122.1 | 134.4 | 125.2 | 127.5 |
| **Combination** | 121.8 | 134.1 | 126.1 | **127.2** |

**Table 6.2 Sedimentation equilibrium analytical ultracentrifugation results for the truncated E2 protein.** The molecular weight in kDa is shown for the three concentrations and two speeds along with the combined and global fits for all results.




**6.11 Sedimentation equilibrium analytical ultracentrifugation of the E2-trunc protein.** The blue dots are the $A_{280}$ recorded with radial distance across the cell after 24 h at 7k r.p.m (B). The red line is the fitted model for a single ideal species with the residuals from fitting the data to this model shown above (A).

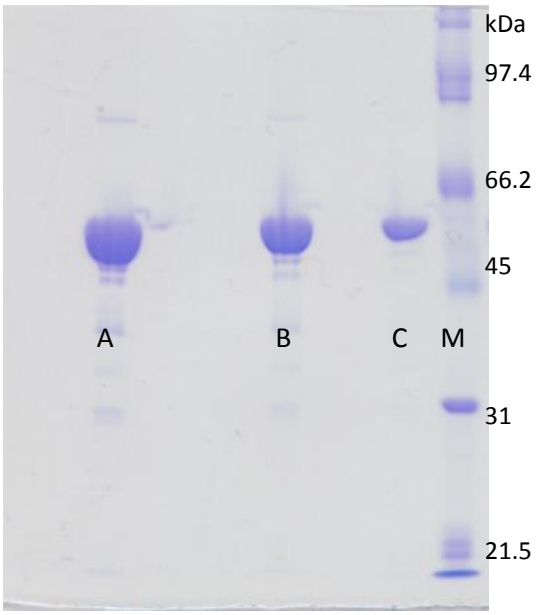



**Figure 6.12 SDS-PAGE of the E2 protein after centrifugation.** The lanes are labelled with the letter that refers to the channel labelled (Table 6.2). The lane labelled M contains the standard protein markers, labelled with their molecular weights (kDa).

The E2-trunc protein molecular weight was determined by fitting the experimental data to a single ideal species model; channel C, the lowest concentration data, fitted the model well (Figure 6.11), and channel B fitted reasonably well. The absorbance in channel A was above 1.5, this is above the linear absorbance range which can lead to the molecular weight being underestimated, and in fact in all cases the molecular weights calculated from channel A data were the lowest. The molecular weight of the E2-trunc protein was calculated to be 127.2 (± 3.0) kDa, the equivalent of 2.8 (± 0.1) monomers. This result strongly suggests that the E2-trunc protein assembles into trimers and that the molecular weight determined by gel filtration was artificially increased by the protein's flexible linkers.

#### *6.3.4.2 Determination of the molecular weight of the truncated E2 catalytic domain*

The data recorded for all three concentrations and speeds fit well to the single ideal species model (Figure 6.13), and therefore all results were used to determine the molecular weight of this protein. The E2cat-trunc protein was calculated to have a molecular weight of 70.2 (± 3.2) kDa, which equates to 2.8 (± 0.1) E2cat-trunc monomers (Table 6.3).

| | **$M_w$ (kDa; data fitted to a single ideal species model)** | | | |
|---|---|---|---|---|
| | **Channel A** | **Channel B** | **Channel C** | **Global fit** |
| **Protein concentration (mg/ml)** | 0.88 | 0.56 | 0.23 | |
| **12k r.p.m 34 h** | 77.4 | 71.2 | 71 | 75.1 |
| **15k r.p.m 30 h** | 73.7 | 69.8 | 65.1 | 71.6 |
| **18k r.p.m. 28 h** | 69.6 | 69.1 | 64.1 | 68.7 |
| **Combination** | 71.8 | 69.5 | 64.6 | **70.2** |

**Table 6.3 Sedimentation equilibrium analytical ultracentrifugation results for the E2cat-trunc protein.** The molecular weight in kDa is shown for the three concentrations and speeds along with the combined and global fit results.

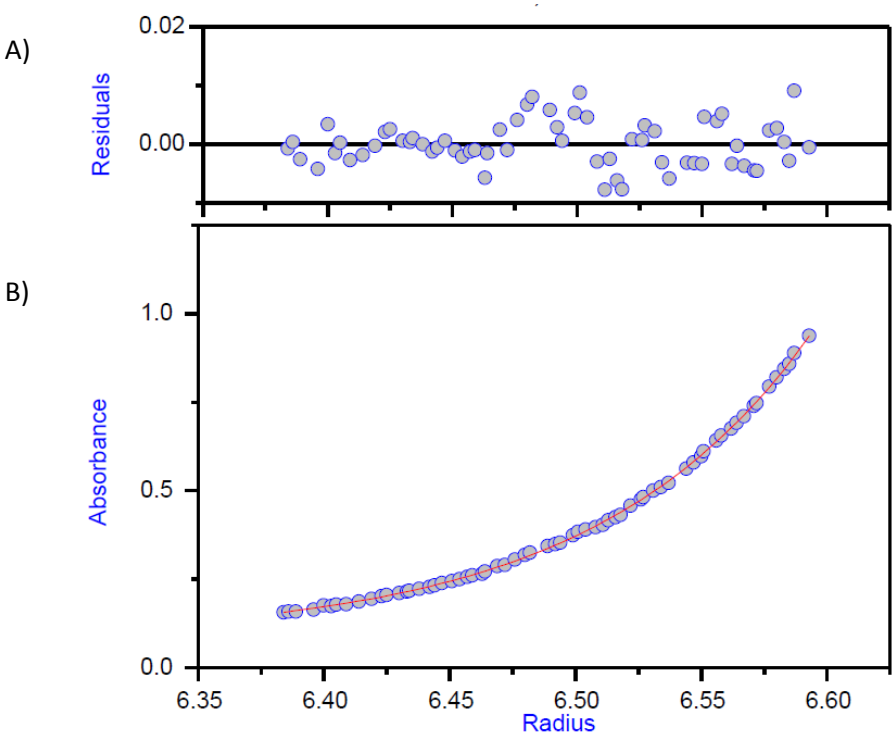



**Figure 6.13 Sedimentation equilibrium analytical ultracentrifugation of the E2cat-trunc protein**. The blue dots are the $A_{280}$ recorded with radial distance across the cell after 30 h at 15k r.p.m (B). The red line is the fitted model for a single ideal species with the residuals from fitting the data to this model shown above (A).

The E2cat-trunc data are more reliable than the E2-trunc, as more of the concentration and speed data sets fitted well to the ideal species model. The molecular weight and multimeric nature of the E2-trunc and E2cat-trunc assemblies, determined by analytical ultracentrifugation, are consistent with one another, indicating that the truncated E2 and E2cat both form trimers.

### 6.3.5 Small-angle X-ray scattering of E2cat-trunc

SAXS was used to determine a low-resolution structure of the E2cat-trunc protein in solution. All experiments were performed on a pure sample of E2cat-trunc (5.8, 4.9, 4.3, 4.0, 3.2, and 0.9 mg/ml) in 50mM Tris, pH 8.8, and 100mM NaCl, the same buffer as was used for the 42-mer E2cat SAXS and all analytical ultracentrifugation experiments.

SAXS confirmed that the E2cat-trunc protein does not form a 42-mer; however, unlike the monodispered E2-trunc and E2cat-trunc trimers detected by analytical ultracentrifugation, a mixture of E2cat-trunc trimers (80%) and hexamers (20%) were seen

by SAXS. A trimer from the non-truncated E2cat 42-mer has been modelled into part of the SAXS *ab initio* bead model (Figure 6.14); the E2cat trimer fits well into this part of the model, and the similarities in size and shape indicate that the E2cat and E2cat-trunc trimers have the same overall structure. Whilst the majority of the E2cat-trunc forms trimers, 20% is hexameric, which may be the result of two trimers interacting with one another. A second trimer cannot be fitted with any certainty into the elongated part of the SAXS model, indicating that this trimer may not have a fixed position relative to the first.

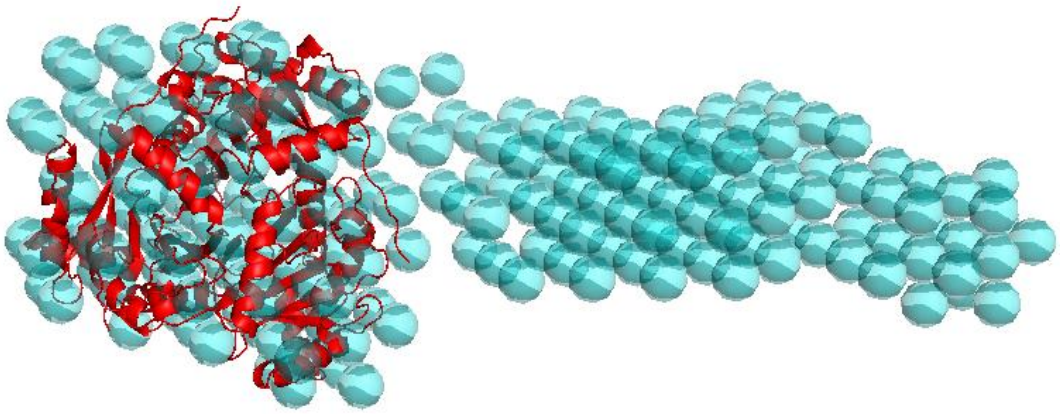


**Figure 6.14 Low resolution *ab initio* bead model of the truncated E2cat protein.** A ribbon diagram of a trimer (red) derived from the *Tp. acidophilum* E2cat 42-mer has been superimposed over the SAXS *ab initio* bead model (cyan).

The SAXS and AUC experiments were performed on pure E2cat-trunc protein in the same buffer conditions; however, higher protein concentrations were used for the SAXS experiments (5.8 - 0.9 mg/ml) compared to the AUC experiments (1.3 - 0.6 mg/ml). The difference in protein concentrations may be the reason why a small proportion of hexamers were detected by SAXS and not AUC. It is not feasible to repeat both sets of experiments using the same concentration range, due to the limitations of each detection method; that is, the AUC centrifuge spectrophotometer is only capable of measuring absorbance <1.2, whereas SAXS requires higher protein concentration as scattering is dependent on protein size and concentration.

### 6.3.6 Truncated E2 catalytic domain crystal trials

Truncating the C-terminus of the E2cat and removing the putative Ile anchor residue produces a polypeptide that assembles into trimers but not a 42-mer. A 42-mer E2cat trimer can be superimposed over the E2cat-trunc SAXS *ab initio* bead model; this indicates that the inability of the E2cat-trunc trimers to form a 42-mer, is not due to a large change in the overall structure of the E2cat trimer. Two possible explanations why the truncated E2 trimers no longer interact are:

1. Shortening the C-terminal helix may disrupt the tertiary structure of the trimer at the trimer interface.
2. The trimer structure may be completely unaffected, but without the essential anchor residue they are unable to interact.

X-ray crystallography has been used to determine the structure of the E2cat-trunc protein, with the aim of elucidating the reason why the truncated trimers do not interact with one another. A secondary aim was to obtain a high-resolution crystal structure of the E2cat trimer, in order to improve the current 4 Å 42-mer model; the E2cat-trunc trimer is smaller (75 kDa) than the 42-mer and therefore may crystallise in a more compact lattice.

#### *6.3.6.1 Initial 96 well plate crystal screen*

The PGA-LM screen (Molecular dimensions, Suffolk, U.K.) resulted in E2cat-trunc crystals in several related conditions containing 5% (w/v) PGA-LM, 0.1M Na cacodylate, pH 6.5, and various molecular weight polyethylene glycols including PEG 400, 550 MME, 2k MME, 4k, 8k and 20k; an example of the crystals that grew in these conditions are shown in Figure 6.15.

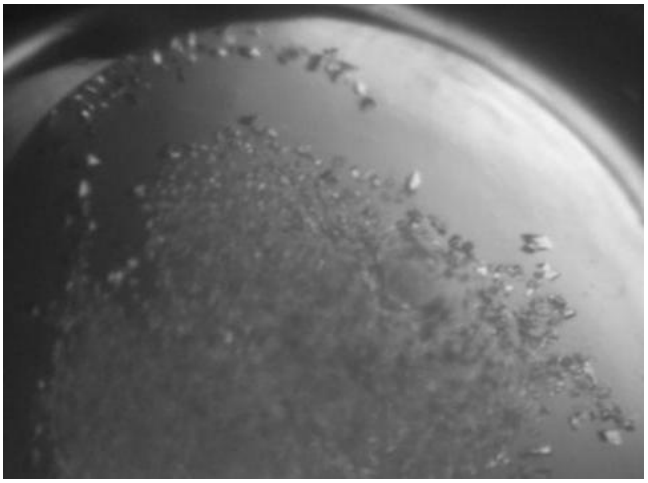


**Figure 6.15 E2cat-trunc crystals in the PGA-LM 96 well crystal screen condition.** The 9mg/ml E2cat-trunc protein was mixed with well solution containing 0.1M Na cacodylate, pH 6.5, 5% (w/v) PGA-LM, and 15% (w/v) PEG 4k. Crystals grew in 5-10 days

#### *6.3.6.2 E2cat-trunc crystal optimisation*

The conditions identified from the PGA-LM screen were reproduced in 24-well hanging-drop trays, and the following components were optimised: storage temperature, PEG molecular weight, the concentration of: PGA, PEG, and protein, and buffer pH. Figure 6.16 shows an example of the crystal morphologies that grew in the optimised condition (8% (w/v) PEG 4k, 3% (w/v) PGA, and 0.1M Na cacodylate, pH 6.5).

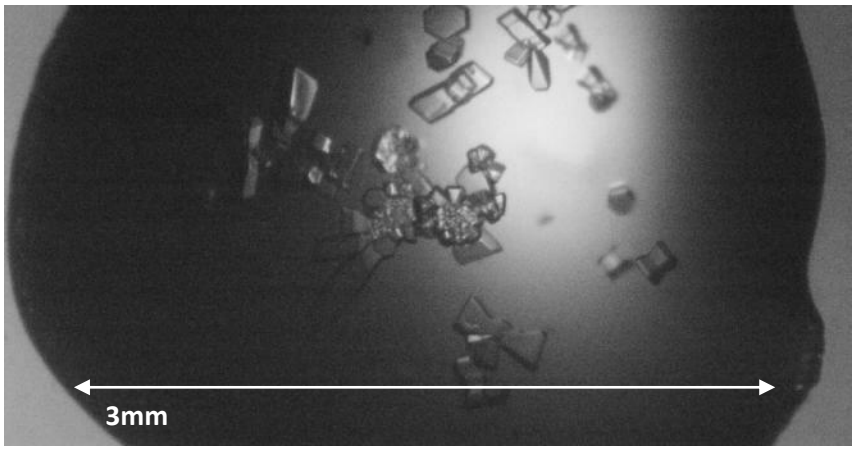



**Figure 6.16 Example of the E2cat-trunc crystals.** The crystals grew in 8% (w/v) PEG 4k, 3% (w/v) PGA, and 0.1M Na cacodylate, pH 6.5, at 16°C.

### 6.3.7 The structure of the truncated *Tp. acidophilum* E2 catalytic domain

#### *6.3.7.1 X-ray diffraction data collection*

Diffraction to 4.1 Å was achieved at the Diamond Light Source (Oxfordshire) with E2cat-trunc crystals (8% (w/v) PEG 4k, 3% (w/v) PGA-LM, and 0.1M Na cacodylate, pH 6.5) that were soaked for 20 s in cryoprotectant (8% (w/v) PEG 4k, 3% (w/v) PGA-LM, 0.1M Na cacodylate, pH 6.5, plus an additional 30% (v/v) glycerol) before being flash frozen. All attempts to improve diffraction resolution by optimising the crystals, freezing conditions and data collections at room temperature, were unsuccessful the crystals either diffracting poorly or showing high levels of mosaicity.

#### *6.3.7.2 Model building and refinement*

The structure of the *Tp. acidophilum* E2cat-trunc protein was determined to a resolution of 4.1Å by molecular replacement using a monomer from the *Tp. acidophilum* E2cat structure (PDB access code: 3RQC).

The asymmetric unit of the E2cat-trunc contains 6 monomers arranged as 2 adjacent trimers (Figure 6.17); the two trimers are, however, in different orientations relative to each other compared to the non-truncated 42-mer E2cat trimers. The polypeptide backbone of the two trimers model was refined to give a final $R_{work}$ of 21.5% and $R_{free}$ of 28.4%.

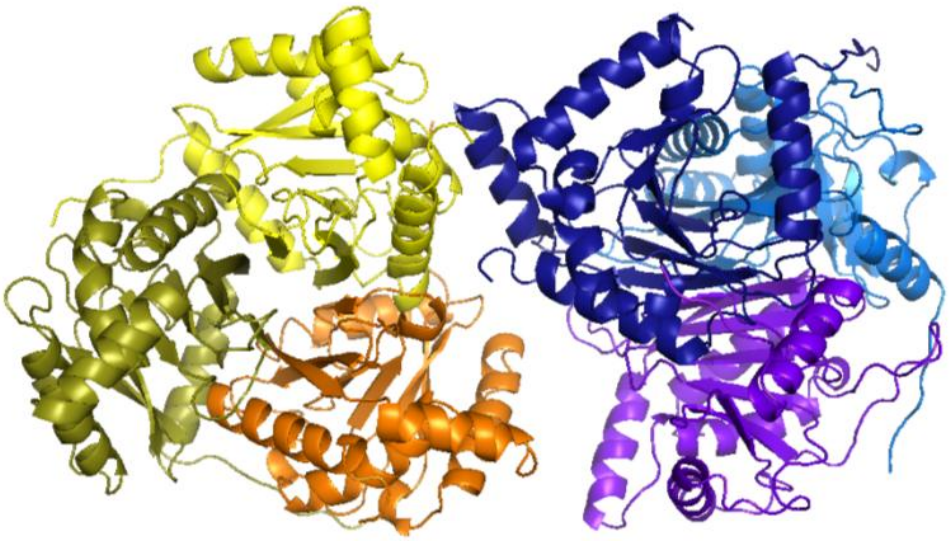
| | | |
|---|---|---|
| **Space Group** | | P32 2 1 |
| **Cell parameters** | a=b, c | 107.2Å, 238.6 Å |
| | α=β, γ | 90°, 120° |
| **Resolution (Å) (last shell)** | | 4.1 (4.17-4.10) |
| **Number of observations** | | 12274 |
| **Number of unique reflections** | | 12902 |
| **Completeness (last shell)** | | 99.2% (97.4) |
| **R merge (%)(last shell)** | | 0.122 (0.551) |
| **Average I/σI (last shell)** | | 13.06 (2.48) |
| **Redundancy** | | 5.3 (5.1) |

**Table 6.4 X-ray data collection statistics from the E2cat-trunc crystal**

| | | |
|---|---|---|
| **Resolution range** | | 92.80- 4.10 Å |
| **Number of protein atoms** | | 10296 |
| **Number of solvent atoms** | | None |
| **Average B value** | | 74.61 $Å^2$ |
| **R-crystal** | | 21.5% |
| **R-free** | | 28.4% |
| **Rms deviations** | **Bonds** | 0.01Å |
| | **Angles** | 1.63° |

**Table 6.5 E2cat-trunc trimer structure refinement statistics**

A)



B)

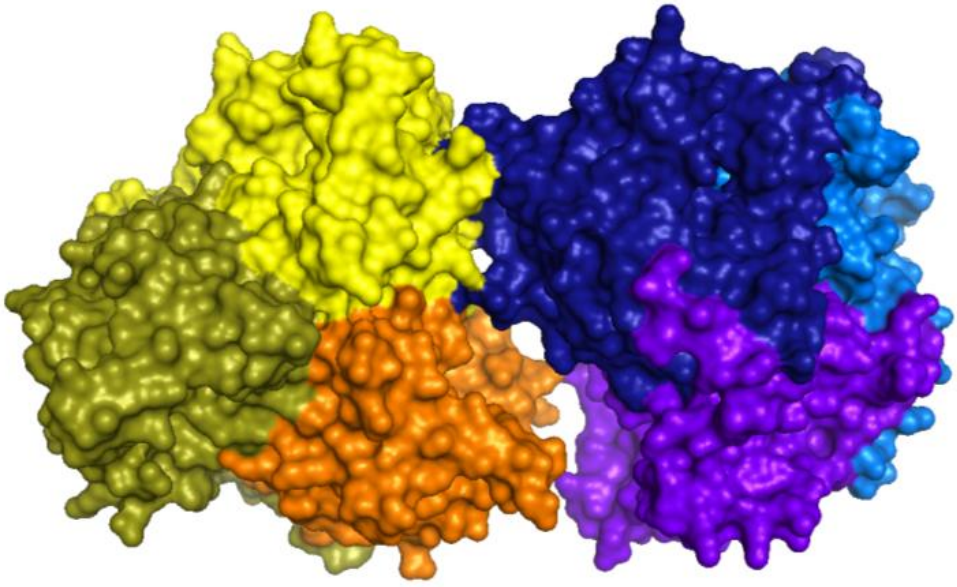


**Figure 6.17 Crystal structure of the truncated E2cat protein to 4.1Å**. Images A and B show a ribbon diagram of the polypeptide backbone structure and space filling model/surface representation, respectively, of the two trimers contained within the asymmetric unit. The image is orientated looking down the three-fold axis of symmetry of the yellow trimer.

### 6.3.8 Analysis of the truncated E2cat crystal structure and comparisons with the 42-mer structure.

#### *6.3.8.1 Comparison of the trimer structure of the truncated and non-truncated E2cat*

The removal of the putative Ile anchor residue, by truncating the C-terminal $3_{10}$ helix, does not affect the overall structure of either the E2 monomer or trimer (Figure 6.18). Monomer A from the trimeric E2cat-trunc and from the E2cat 42-mer have a Cα RMSD value of 0.86Å (214/219 residues); furthermore, two intra-trimer monomers (monomer A and B) taken from an E2cat-trunc trimer, and E2cat trimer have a Cα RMSD value of 0.71Å (214/219) and 0.77Å (217/222), respectively. The E2cat and E2cat-trunc trimers appear to be essentially identical; truncating the C-terminal helix does not change the tertiary structure of the protein. Therefore, it can be concluded that the C-terminal anchor residue is crucial in the formation of the inter-trimer interactions.

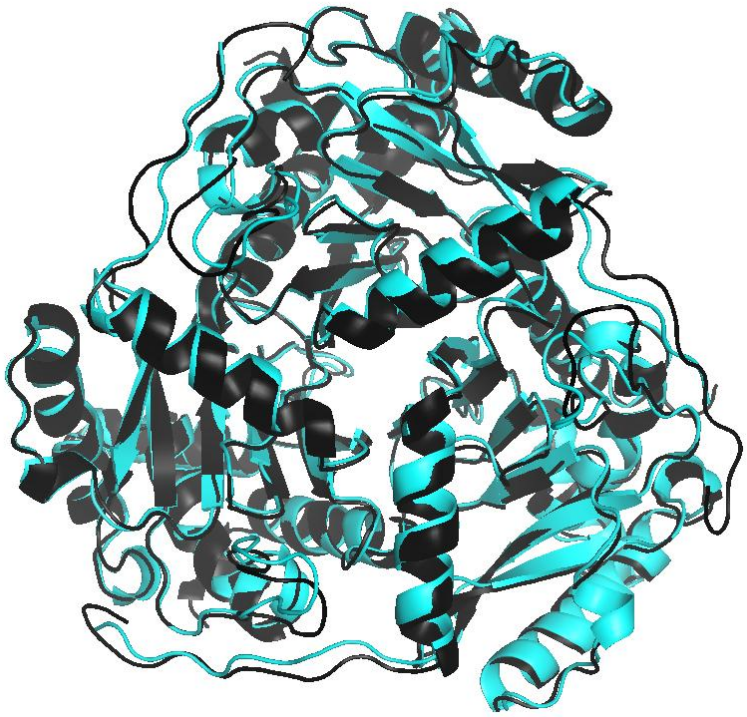


**Figure 6.18 E2-trunc trimer superimposed over an E2 trimer from the 42-mer crystal structure.** The E2-trunc trimer structure is shown in blue and the 42-mer trimer is shown in black.

### *6.3.8.2 Orientation of the two trimers within the asymmetric unit*

The asymmetric unit of the E2cat-trunc crystal structure contains two trimers, but these are orientated differently relative to each other, compared to the position of two 42-mer E2cat trimers (Figure 6.19). The C-terminal anchor residue appears to be crucial in the formation of the 42-mer trimer-trimer interactions; without this region, the trimers do not assemble correctly, despite the residual of the trimer-trimer interface remaining unaltered.

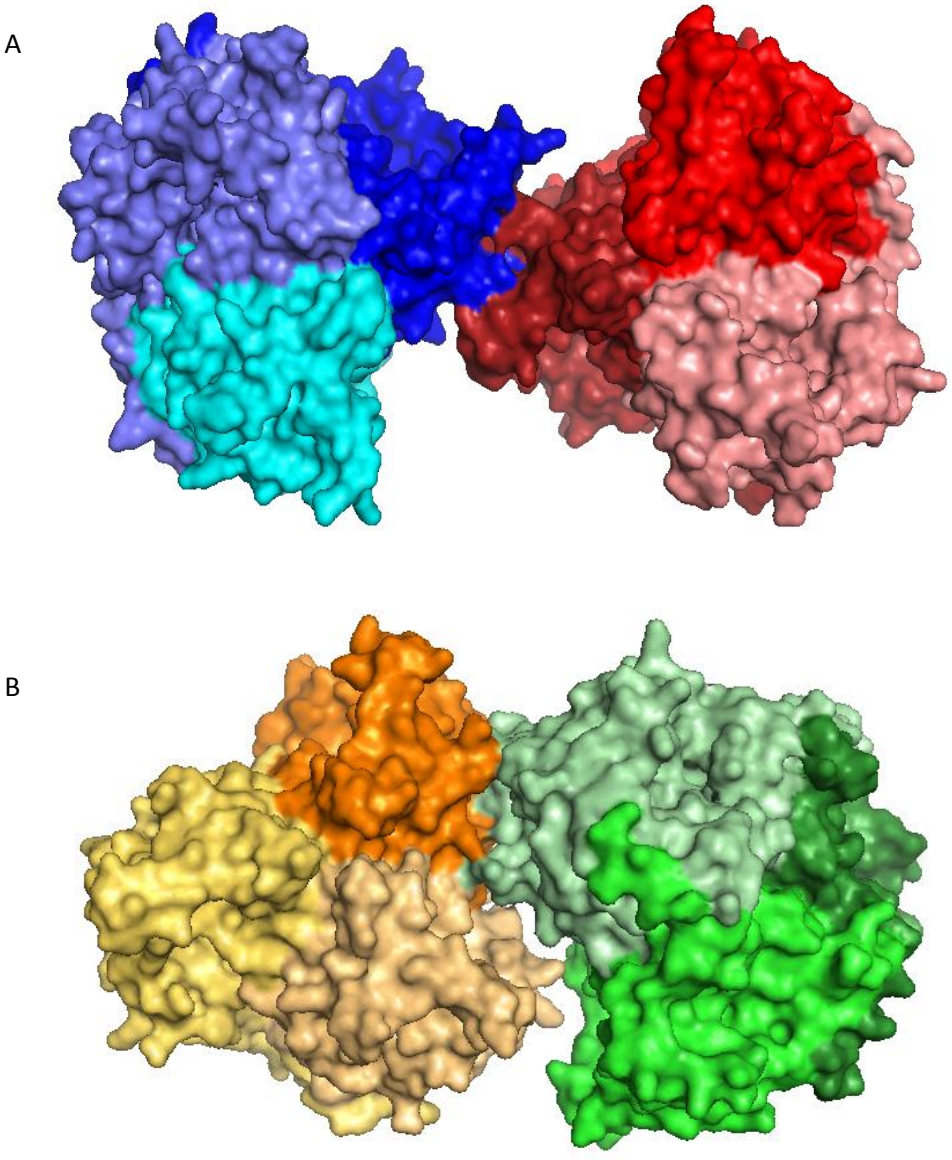



**Figure 6.19 The relative orientations of two trimers in the 42mer E2cat structure (A), and the trimeric E2cat-trunc (B).** Space filling diagrams of two E2cat trimers shown in blue and red and two E2cat-trunc trimers shown in orange and green.

## 6.4 Conclusion

The E2 anchor residue has been predicted to play a key role in the formation of trimer-trimer interactions. Two possible *Tp. acidophilum* E2 anchor residues were identified by sequence alignments with other E2 sequences with known structures, they were E2-Ile 396 and E2-Ile 397; based on the *Tp. acidophilum* E2cat structure, Ile 397 would seem the more likely candidate, however this has not been unequivocally confirmed.

The conservative mutation of the E2cat Ile 397, to a valine or a leucine, resulted in insoluble protein; whereas, removal of both isoleucines, by the insertion of a premature stop codon before the final 5 C-terminal residues (IIYEI), resulted in soluble recombinant E2 and E2cat proteins. The molecular weights of the E2-trunc and E2cat-trunc protein determined by equilibrium analytical ultracentrifugation (AUC) equated to a trimeric assembly, rather than a 42-mer. Unlike the monodisperse E2cat-trunc assembly seen by AUC a mixture of trimers (80%) and hexamers (20%) were seen by small-angle X-ray scattering (SAXS). The variation between the two methods could be explained by different protein concentrations: SAXS requires a much higher concentration than AUC, and therefore any weak interactions that result in trimer-dimers are only seen at the high protein concentrations used in the SAXS experiments.

The structure of the E2cat-trunc trimer was determined by X-ray crystallography to 4.1 Å resolution. The structure of the trimeric E2cat-trunc trimer and the 42-mer E2cat trimer are the same; removing the C-terminal anchor residue has no effect on the structure of the trimer, but it completely disrupts the 42-mer trimer-trimer interactions. The results discussed in this chapter are consistent with work by Peng et al. (2012) on the dodecahedral *G. stearothermophilus* E2; in this case, the whole C-terminal helix (DPELLLMEA) was removed, resulting in a trimeric E2 catalytic domain protein.

The C-terminal anchor residue region has been shown to be crucial in the formation of a 42-mer; the angle at which the trimers interact with one another may be influenced by how the anchor residue fits into the hydrophobic pocket on the adjacent trimer. As site-directed mutagenesis of the putative anchor residue was unsuccessful, in that the protein was expressed in an insoluble form, the hydrophobic pocket residues could be mutated to investigate how the anchor residue interactions affect the angle at which the trimers interact. Further experiments are also needed to confirm if E2-Ile 397 rather than Ile 396 is the *Tp. acidophilum* E2 anchor residue.

# Chapter 7: Characterisation of the trimeric E2 and OADHC

## 7.1 Introduction

A trimeric version of the *Thermoplasma acidophilum* E2 catalytic domain, and full length E2, along with its N-terminal PSBD and lipoyl domain, have been produced by removing the inter-trimer anchor residue, a C-terminal Ile that extends from one trimer into a hydrophobic pocket on the adjacent trimer and *vice versa*. The removal of this key residue, by truncating the C-terminus of the catalytic domain, prevents the E2 trimers associating into a 42-mer. There are three acyltransferase active sites per trimer, situated in a tunnel at each of the monomer-monomer interfaces; the key active site residues are provided by both monomers: one monomer provides a His and an Asp, and the other a Thr (Figure 7.1).

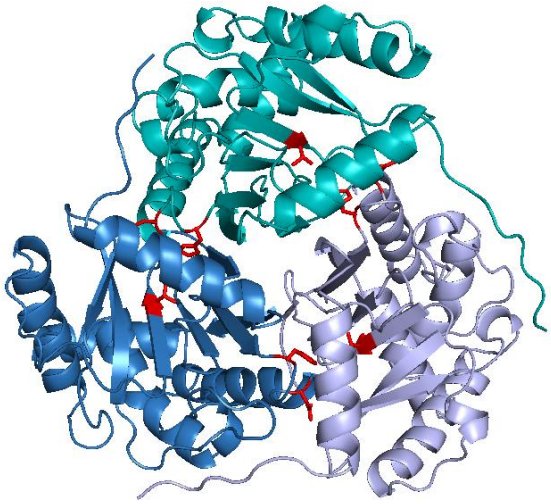


**Figure 7.1 Crystal structure of the trimeric E2 catalytic domain.** The three monomers are coloured in three shades of blue, and the active site residues (Thr, His, and Asp) are highlighted in red. There are three active sites per trimer, with one monomer contributing the His and Asp and a second monomer contributing a Thr.

The structure of the trimeric E2cat, determined in Chapter 6, superimposes over the 42-mer E2cat trimer, this suggests that removing the E2cat's C-terminal anchor residue does not alter the structure of the trimer, only its ability to interact with other trimers and assemble into a 42-mer. If the structure of the trimer remains identical, then it can be hypothesised that it may retain acyltransferase activity. The N-terminal region remains unaltered by this mutation, and therefore the full-length E2 has the PSBD and lipoyl domain required for OADHC assembly and activity. The aim of this chapter was to characterise the trimeric *Tp. acidophilum* E2cat and the trimeric full-length E2, and to investigate if it is possible to produce a functional trimeric OADHC.

## 7.2 Methods

### 7.2.1 Expression and purification of truncated E2 and E2 catalytic domain proteins

The truncated E2cat (E2cat-trunc) was expressed as described in Chapter 2.3.1. The truncated E2 (E2-trunc) was expressed and lipoylated as described in Chapter 3.2.2.

The E2-trunc was purified by the method described in Chapter 2.3.3, and resulted in protein that was approximately 90% pure. The E2cat-trunc protein was purified by the 4-step purification method established for the purification of the E2cat protein in Chapter 4.2.2; this method was also used when the E2-trunc protein was required to be >95% pure and in a buffer not containing imidazole.

### 7.2.2 E2 dihydrolipoamide transacetylase assay

The E2 assay is described in Chapter 2.5.2. Unless otherwise stated, the non-lipoylated E2-trunc was utilised for all E2 activity assays, rather than the lipoylated form of the enzyme, as previous results suggest that the enzyme-bound lipoic acid competes with the exogenous DHlip (Chapter 3.3.2.2).

### 7.2.3 Assembly and OADHC activity of a trimeric complex

#### *7.2.3.1 Expression and purification of the E1 and E3 enzyme*

The E1 and E3 enzymes were expressed as described in Chapter 2.3.1 and purified following the method in Chapter 2.3.3. The E1, E2-trunc, and E3 enzymes were dialysed overnight, at 4°C, into 50mM Tris, pH 8.5, 100mM NaCl and 50mM imidazole.

#### *7.2.3.2 2-oxoacid dehydrogenase complex assembly and activity assay*

The OADHC assays were performed as described in Chapter 2.5.4. The concentrations of the purified enzymes were determined by measuring the absorbance at 280nm or using the Bradford assay, depending on the purity of the enzyme (Chapter 2.4). The protein concentration was then converted to a molar concentration of the active protein units: $E1\alpha_2\beta_2$, $E2\alpha$, and $E3\alpha_2$, using the molar absorbance coefficient predicted from the protein sequences (Gasteiger et al., 2003). The required molar ratio of $E1\alpha_2\beta_2$, $E2\alpha$, and $E3\alpha_2$ were mixed in assay buffer in the cuvette (50mM potassium phosphate pH7.0, 2.5mM

$NAD^+$, 1mM MgCl2, and 0.2mM TPP) and incubated at 55°C for 10 min before the reaction was started with the addition of 0.13mM CoA, 2.6mM cysteine/HCl and the 2-oxoacid substrate (e.g. 2mM 3-methyl-2-oxopentanoic acid).

The $K_m$ and $V_{max}$ values of the E2 or E2-trunc OADHC with respect to 3-methyl-2-oxopentanoic acid were determined under comparable conditions by mixing the $E1\alpha_2\beta_2$:E2α:$E3\alpha_2$ in a stoichiometry of 1:1:0.5.

#### *7.2.3.3 Gel filtration of the assembled trimeric OADHC*

The $E1\alpha_2\beta_2$, E2α-trunc, and $E3\alpha_2$ were mixed in a ratio of 1:1:1 and heated at 55°C for 10 min without dilution; any precipitate was removed by centrifugation at 16 000 x g and filtration through a 0.45µm filter. The soluble OADHC sample was then loaded onto a Superdex 200, 10-300GL (GE Healthcare) column equilibrated with 25mM Tris-HCl, pH 8.5, 100mM NaCl and 25mM imidazole. The column eluent was collected in 0.5ml fractions; E1, E3 and OADHC activities were measured for each fraction, along with the determination of the protein concentration by the Bradford assay (Chapter 2.4.1).

### 7.2.4 Characterisation of the temperature optima and thermal inactivation profiles of the truncated E2 and OADHC

#### *7.2.4.1 Thermal inactivation assays*

The thermal inactivation assays were performed as described in Chapter 3.2.6. The E2-trunc was incubated in 50mM Tris-HCl, pH 8.8, and 100mM NaCl, and the E2cat-trunc was incubated in 50mM Tris-HCl, pH8.8. The OADHC components were mixed in the required stoichiometry to produce a stock solution; all aliquots to be heated were taken from this solution to ensure a constant stoichiometry.

#### *7.2.4.2 Temperature optima*

The temperature optima assays were performed as described in Chapter 3.2.7. A stock of assembled OADHC was made by mixing the component enzymes in a stoichiometry of 1:1:0.5 ($E1\alpha_2\beta_2$:E2:$E3\alpha_2$) and incubated in complex assay buffer at 55°C for 10 min, the OADHC stock was then cooled on ice and stored at 4°C. The reactions were started with the addition of enzyme rather than substrate.

## 7.3 Results and discussion

### 7.3.1 Characterisation of the truncated E2 and E2cat dihydrolipoamide transacetylase activity

#### *7.3.1.1 The dihydrolipoamide transacetylase activity of the E2-trunc and E2cat-trunc proteins*

The enzymatic activities of the truncated E2 and E2cat were compared to that of the non-truncated versions, to determine if the truncation and prevention of 42-mer assembly influenced the enzymes' catalytic ability. The recorded specific activities (Table 6.1) show that both the E2-trunc and E2cat-trunc are capable of catalysing the acyltransferase reaction, albeit at reduced rate; compared to the non-truncated versions, the specific activity of the E2-trunc had been reduced to 78% and the E2cat-trunc to 36%. Thus, truncating the E2 and preventing assembly does not abolish activity but does reduce the enzymes' efficiency.

| Enzyme | Specific activity (U/mg) | $k_{cat}$ ($min^{-1}$) |
|---|---|---|
| **E2** | 1.06 ± 0.05 | 48.9 ±2.3 |
| **E2-trunc** | 0.83± 0.01 | 37.8 ± 0.2 |
| **E2cat** | 1.63 ±0.13 | 42.8 ± 3.4 |
| **E2cat-trunc** | 0.58 ± 0.08 | 14.5 ± 2.0 |

**Table 7.1 Specific activity and $k_{cat}$ values for the truncated and non-truncated E2 and E2cat.** The specific activity is shown in U/mg of E2, where 1 Unit is described as the conversion of 1 µmol of acetyl-CoA to 1 µmol of acetyl-dihydrolipoamide per minute.

#### *7.3.1.2 Thermal inactivation of the truncated E2 catalytic domain*

The thermostability of the truncated proteins can be measured as both the E2-trunc and E2cat-trunc are catalytically active. The thermostability assays of the truncated and the non-truncated 42-mer E2 and E2cat proteins (Chapters 3.3.2 and 4.3.5) were conducted under the same conditions; therefore, the half-lives can be directly compared to determine if preventing the trimers associating into a 42-mer affects the proteins' thermostability (Figure 7.4).

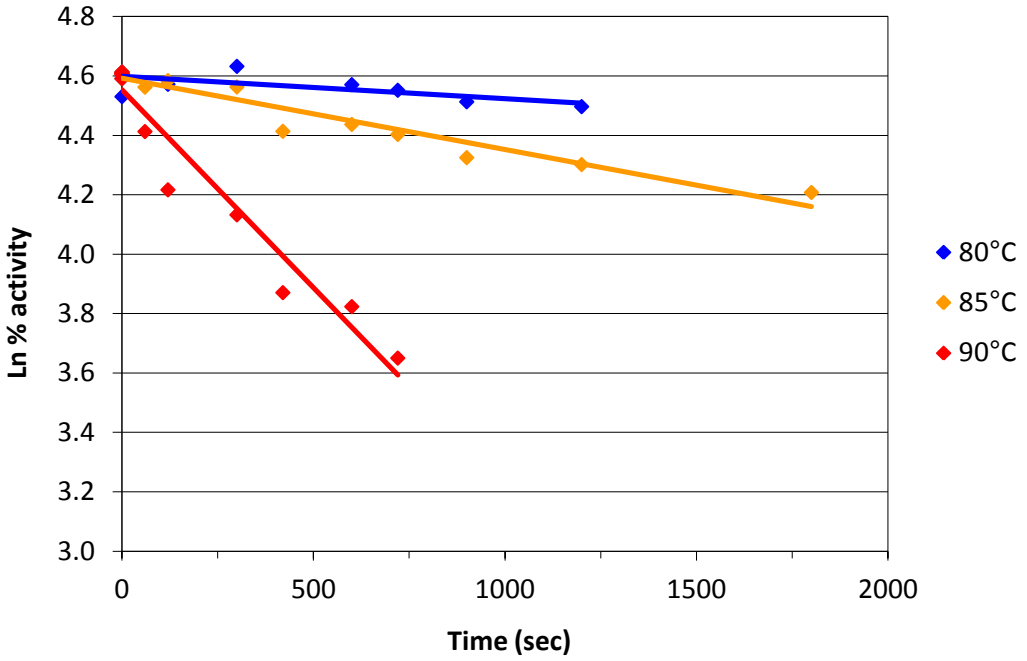



**Figure 7.2 Thermal inactivation profile of the truncated E2**. The truncated E2 protein was incubated in 50mM Tris-HCl, pH 8.8, 100mM NaCl

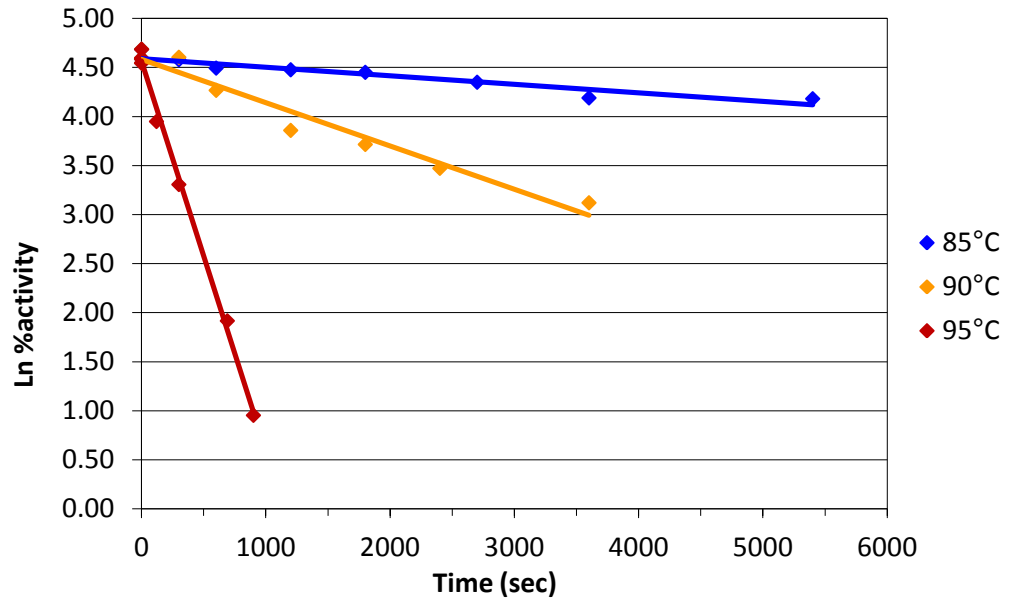



**Figure 7.3 Thermal inactivation profile of the truncated E2cat.** The truncated E2cat protein was incubated in 50mM Tris-HCl, pH 8.8

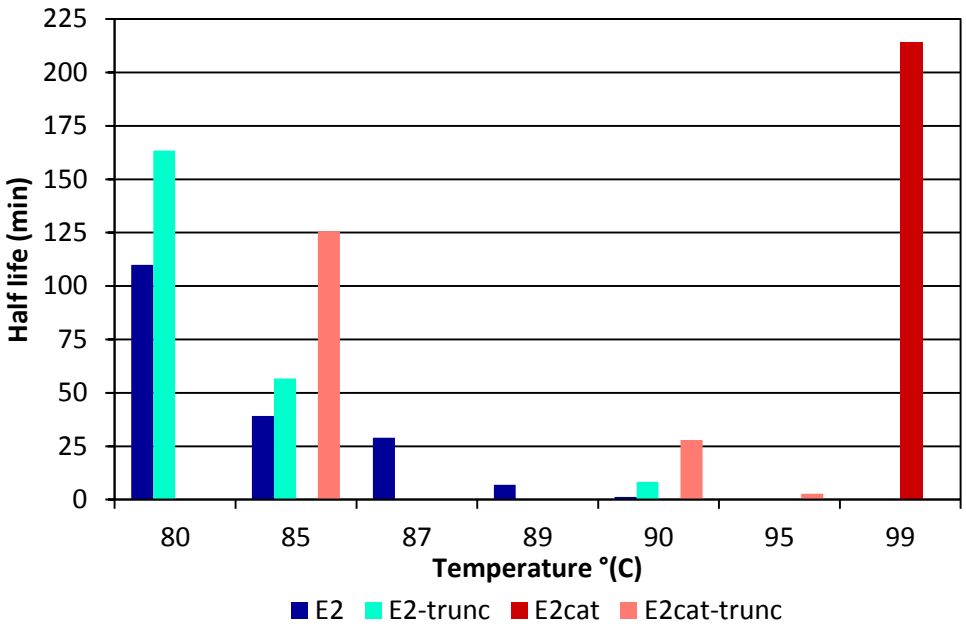



**Figure 7.4 Bar chart comparing the half-lives of the trimeric E2-trunc, and E2cat-trunc with the non-truncated42-mer E2 and E2cat**

The trimeric E2cat-trunc is significantly less thermostable than the non-truncated 42-mer E2cat; the E2cat has a half-life of 215 min at 99°C, whereas, the E2cat-trunc has a half-life of just 3 min at 95°C (Figure 7.4). There is less of a difference between the thermostability of the E2 trimer and 42-mer; at 85°C both versions have a similar half-life of 56 and 39 min, respectively (Figure 7.4). This data set implies that the trimeric E2-trunc is slightly more thermostable than the 42-mer E2; however, further replicates of these results would be required to determine if there is a statistically significant difference between the thermostabilities.

Truncating the E2cat protein reduces its thermostability to a greater extent than truncating the E2 protein. Interestingly, the presence or absence of the flexible linker regions has more of an effect on the thermostability of the E2, than preventing the trimers assembling into a 42-mer.

### 7.3.2 Truncated E2 and OADHC activity

The trimeric E2 is capable of catalysing the dihydrolipoamide acyltransferase reaction, albeit at a reduced rate compared to the fully-assembled 42-mer E2. The removal of the C-terminal anchor residue is unlikely to influence the peripheral subunit binding domain or lipoyl domain; therefore, the E2-trunc should be capable of binding the E1 and E3 enzymes, and forming an OADHC.

#### *7.3.2.1 Trimeric and functional OADHC detection*

The specific activity values shown in Table 7.2 demonstrate that when lipoylated E2-trunc is mixed with E1 and E3, it is capable of catalysing the OADHC reaction. The order of substrate preference is the same as the pattern seen for the E1 (Chapter 3.3.1.3), with the highest activity seen with 3-methyl-2-oxopentanoic acid, low activity with pyruvate and no activity with 2-oxoglutarate.

| | **Specific activity** | |
|---|---|---|
| **Substrate** | **U/mg complex** | **U/mg E2** |
| **3-methyl-2-oxopentanoic acid** | 0.013 ± 0.001 | 0.068 ± 0.003 |
| **4-methyl-2-oxopentanoic acid** | 0.010 ± 0.001 | 0.050 ± 0.005 |
| **Pyruvate** | 0.006 ± 0.002 | 0.030 ± 0.009 |

**Table 7.2 Specific activity of the trimeric OADHC with a variety of 2-oxoacid substrates.** The components were mixed in a stoichiometry of 1:1:0.5 ($E1\alpha_2\beta_2$:E2α:$E3\alpha_2$) and assayed with 1mM 2-oxoacid. The specific activity is shown in U/mg of E2 and U/mg of complex, where 1 Unit is described as the conversion of 1 µmol $NAD^+$ to NADH per minute.

#### *7.3.2.2 Comparison of kinetic parameters of the trimer and 42-mer complexes*

The $K_m$ and $V_{max}$ values were determined under comparable conditions for the 42-mer and trimeric OADHCs; the component enzymes were mixed in a stoichiometry of 1:1:0.5 ($E1\alpha_2\beta_2$:E2α:$E3\alpha_2$) (Table 7.3 and Figures 7.5-7.8).

| **Complex structure** | $K_m$ | $V_{max}$ | | $V_{max}/K_m$ |
|---|---|---|---|---|
| | **mM** | **U/mg Complex** | **U/mg of E2** | **U/mg of E2 / mM** |
| **42-mer** | 0.306 ± 0.011 | 0.173 ± 0.002 | 0.922 ± 0.009 | 3.02 ± 0.11 |
| **Trimer** | 0.122 ± 0.002 | 0.080 ± 0.001 | 0.430 ± 0.002 | 3.51 ±0.07 |

**Table 7.3 Comparison of the $K_m$ and $V_{max}$ values for the trimeric complex (E2-trunc) and the 42-mer complex (non-truncated E2) when utilising 3-methyl-2-oxopentaonic acid.** The component enzymes were mixed in the cuvette in a stoichiometry of 1:1:0.5 (E1α2β2:E2α:E3α2). The specific activity is shown in U/mg of E2 and U/mg of complex, where 1 Unit is described as the conversion of 1 µmol NAD+ to NADH per minute.

The $V_{max}$ and $K_m$ values determined for the trimeric OADHC are 54% and 60% lower than that of the 42-mer OADHC, respectively. The $V_{max}/K_m$ values for the two complexes are similar, therefore implying, but not explicitly stating, that the catalytic efficiencies of the two complexes are similar (Eisenthal et al., 2007). From these experiments it is not known why the kinetic parameters of the 42-mer and trimeric complexes are different. The E1 enzyme is usually the rate limiting step in the OADHC catalysed reaction. Therefore, the $V_{max}$ and $K_m$ of the complex will depend upon the amount of E1 bound to the E2; preventing the trimers from assembling into a 42-mer may influence the amount of E1 bound to each E2 trimer.

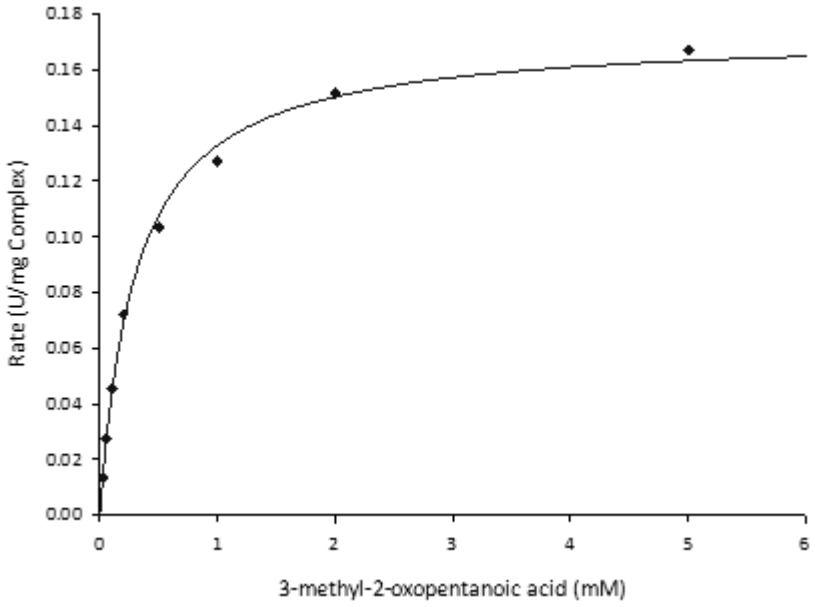



**Figure 7.5 Michaelis-Menten plot of 42-mer OADHC activity with 3-methyl-2-oxopentanoic acid.** The 42-mer OADHC is composed of E1:E2:E3 in a stoichiometry of 1:1:0.5

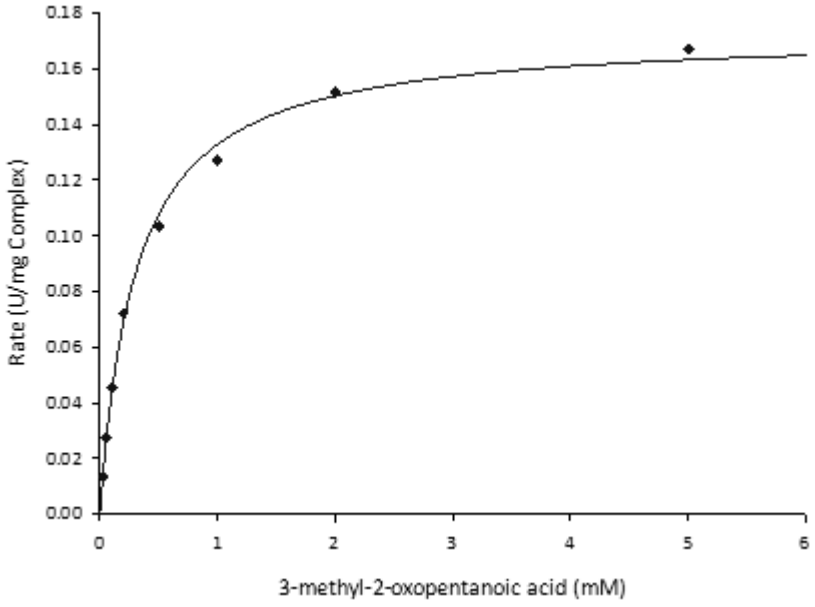



**Figure 7.6 Hanes-Woolf plot of 42-mer OADHC activity with 3-methyl-2-oxopentanoic acid.** The 42-mer OADHC is composed of E1:E2:E3 in a stoichiometry of 1:1:0.5

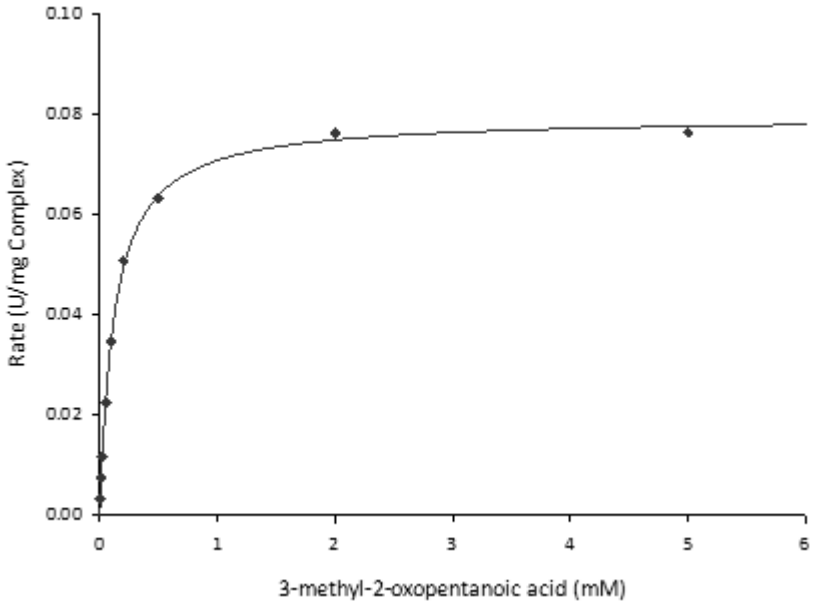



**Figure 7.7 Michaelis-Menten plot of trimeric OADHC activity with 3-methyl-2-oxopentanoic acid.** The trimeric OADHC is composed of E1:E2-trunc:E3 in a stoichiometry of 1:1:0.5

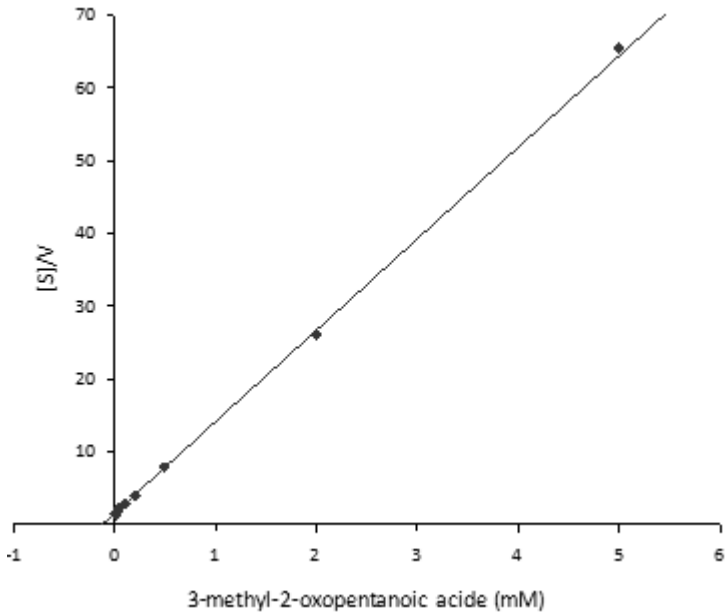



**Figure 7.8 Hanes-Woolf plot of trimeric OADHC activity with 3-methyl-2-oxopentanoic acid.** The trimeric OADHC is composed of E1:E2-trunc:E3 in a stoichiometry of 1:1:0.5

### 7.3.3 Assembly of the trimeric OADHC

#### *7.3.3.1 Stoichiometry of the trimeric E2 complex*

The OADHC activity was measured with different stoichiometries of E1 and E3, in order to determine if varying the ratio of the E1 and E3 to the trimeric and 42-mer E2 cores affects OADHC activity in a similar pattern. Optimal activity for the trimeric OADHC was achieved with the stoichiometry of 1:1:1 ($E1\alpha_2\beta_2$:E2-trunc$\alpha$:$E3\alpha_2$) (Figure 7.9), compared to the optimal stoichiometry for the 42-mer OADHC of 1.75:1:0.5 ($E1\alpha_2\beta_2$:E2$\alpha$:$E3\alpha_2$) (Chapter3.3.5).

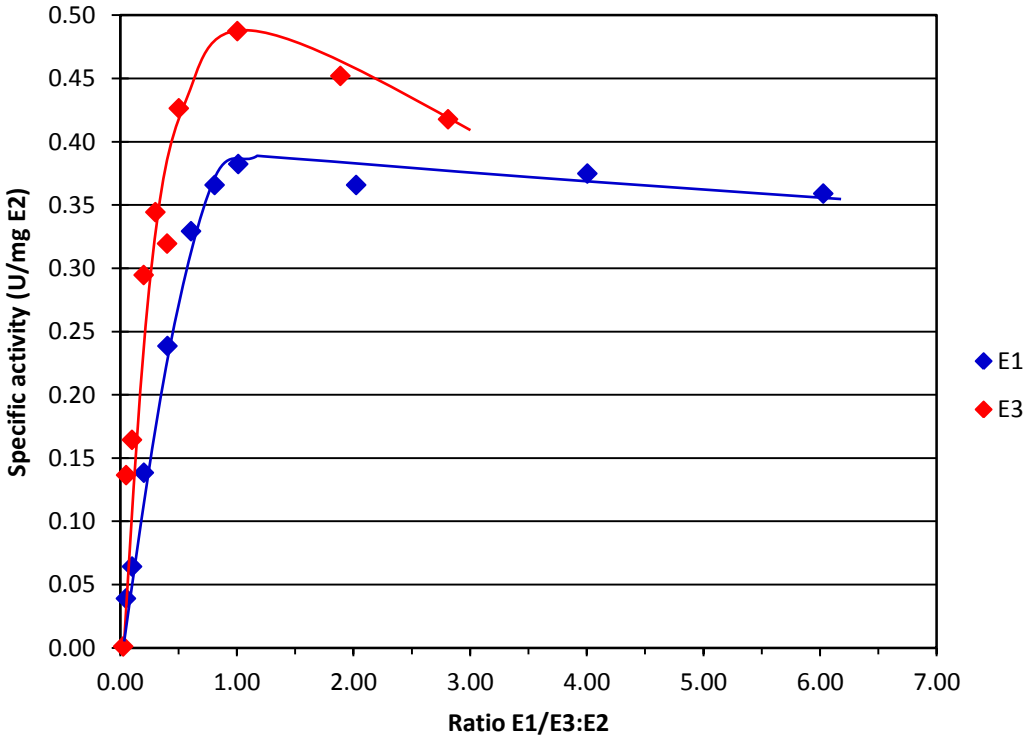



**Figure 7.9 Specific activity of the trimeric complex as a function of the ratios of E1 and E3 to E2.** The ratio of $E1\alpha_2\beta_2$ and $E3\alpha_2$ were varied independently with respect to the concentration of E2$\alpha$. The E1:E2 ratio was fixed at 1:1 when the amount of E3 was varied, and the ratio of E2:E3 was fixed at 1:0.5 when the amount of E1 was varied.

Figures 7.10 and 7.11 compare the relative OADHC activities when the ratios of E1 and E3 were varied with respect to 42-mer E2 and trimeric E2-trunc.

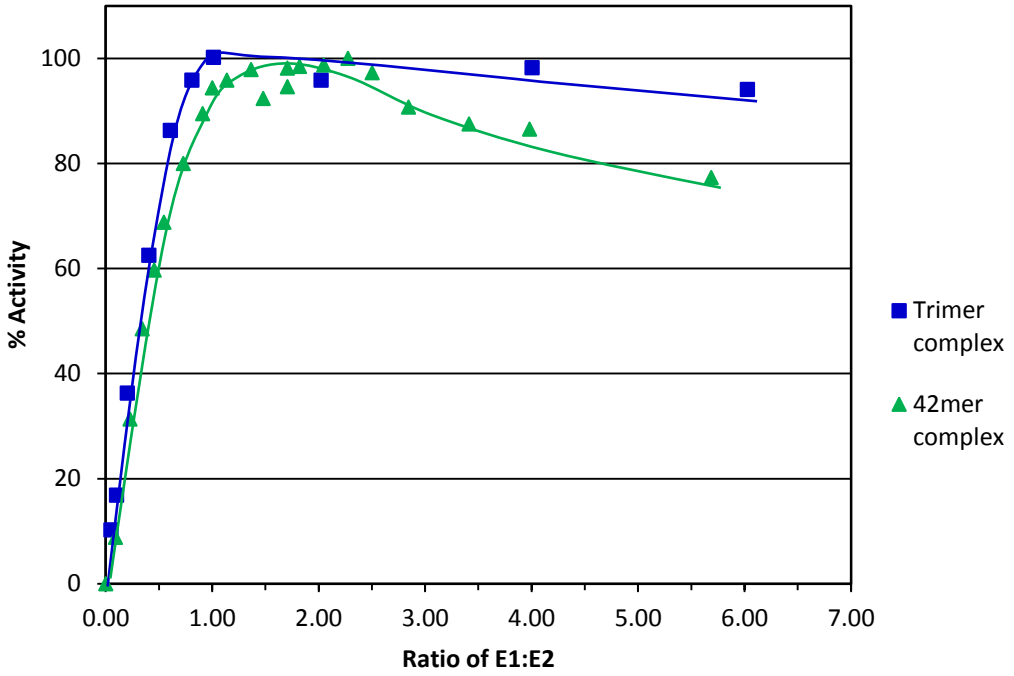



**Figure 7.10 A comparison of the effect of changing the E1:E2 stoichiometry on the OADHC activity of the 42-mer and trimeric complexes.** The amount of $E1\alpha_2\beta_2$ was varied with respected to the amount of 42-mer E2 or the trimeric E2-trunc; the ratio of 42-mer $E2\alpha$: $E3\alpha_2$ was fixed at 1:1, and the ratio of trimeric E2-trunc:$E3\alpha_2$ was fixed at 1:0.5

The ratio of the 42-mer E2:E3 was fixed at 1:1, whereas the trimeric E2-trunc:E3 ratio was fixed at 1:0.5. There appears to be less competition for E1 to bind to the E2-trunc PSBD from E3, than in the non-truncated E2 experiments; this variation may explain why optimal OADHC activity was achieved at a ratio of 1:1 E1:E2-trunc and 1.5:1 E1:E2.

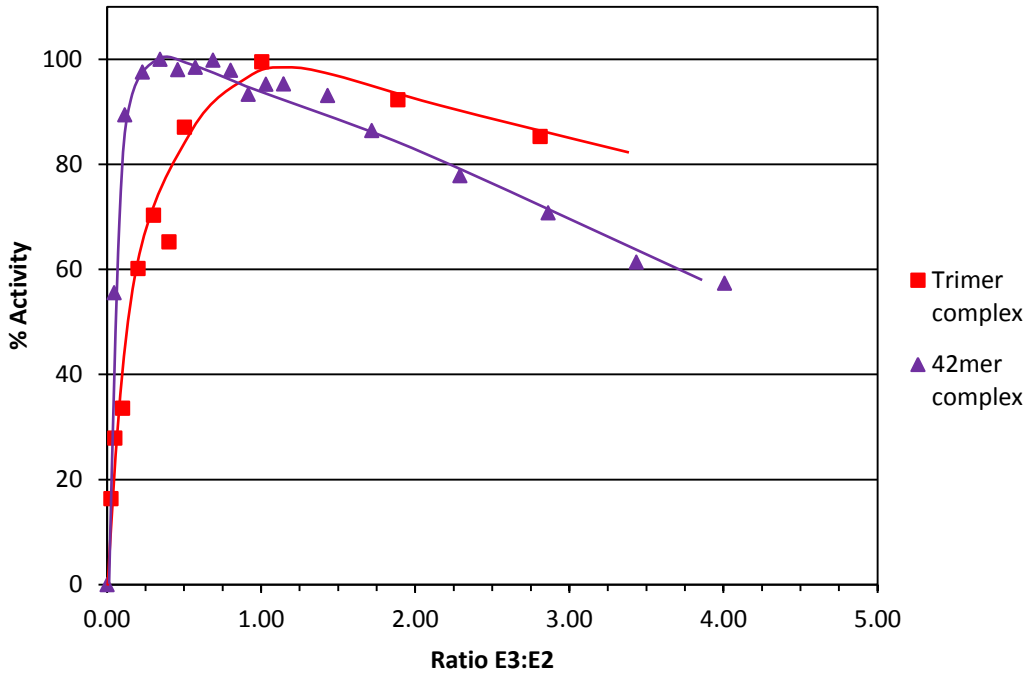



**Figure 7.11 A comparison of the effect of changing the stoichiometry of E3:E2 on the OADHC activity of the 42-mer and trimeric complexes.** The amount of E3 was varied with respect to 42-mer E2 and trimeric E2-trunc; the ratio of E1$\alpha_2\beta_2$:E2/E2-trunc was fixed at 1:1.

To achieve optimal activity, a higher ratio of E3:E2 was required for the trimeric complex than for the 42-mer complex. In the case of the 42-mer OADHC, optimal activity was achieved with an excess of E1:E3. The E1 enzyme is probably the rate limiting step in the reaction catalysed by the 42-mer OADHC, which may be capable of channelling acyl groups around the complex faster than the E1 can catalyse the decarboxylation reaction (Danson et al., 1978). The reduction in the number of E3 enzymes per complex does not compromise overall activity, as the E3 enzyme is the fastest of the three component enzymes, and therefore is capable of regenerating the lipoic moiety as quickly as the E1 acylates it. The trimeric OADHC requires a higher ratio of E3:E1 than the 42-mer complex; this complex is most likely not capable of channelling substrates between trimers, and so consequently each trimer must have at least one E1 and one E3 bound. The E3 binds less tightly to the PSBD than the E1 (gel filtration experiments in Figures 7.13 and 7.14); therefore, a greater concentration of E3 is required to displace an E1 and occupy at least one PSBD per trimer.

### *7.3.3.2 Gel filtration of the trimeric complex*

Size exclusion gel filtration chromatography was performed to examine the stability of the assembled trimeric complex. The purified complex components in 25mM Tris-HCl, pH 8.5, 100mM NaCl, and 25mM imidazole were mixed in a molar stoichiometry of 1:1:1 ($E1\alpha_2\beta_2$:E2:$E3\alpha_2$) and incubated at 55°C for 10 min to allow assembly. The gel filtration column was equilibrated with 25mM Tris-HCl, pH8.5, 100mM NaCl, and 25mM imidazole.

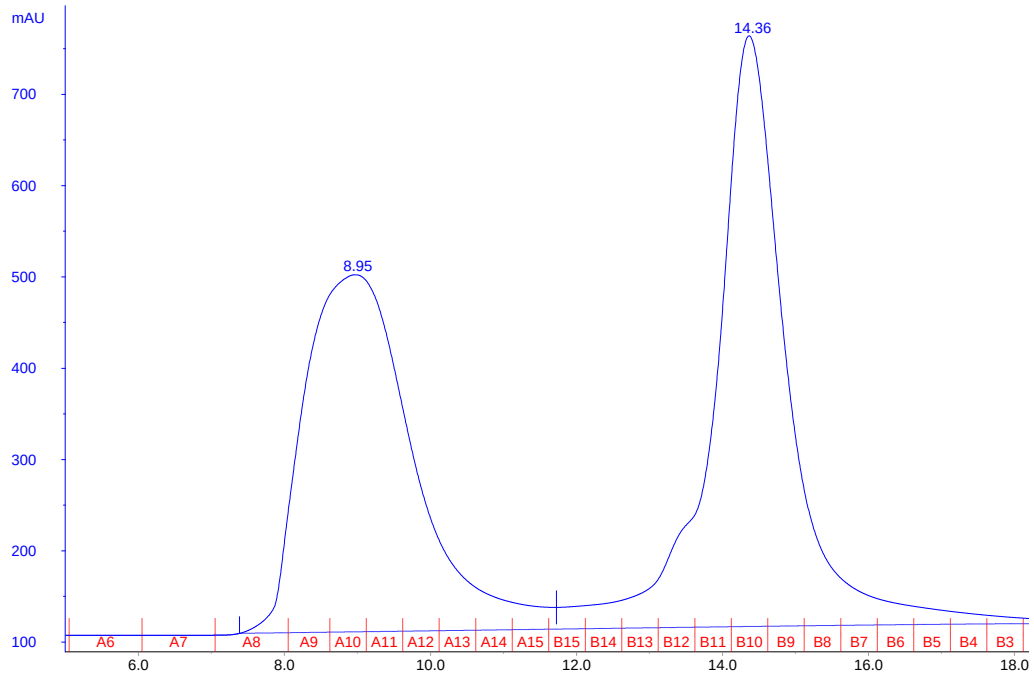



**Figure 7.12 Gel filtration trace of the trimeric OADHC.** The $A_{280}$ readings are shown in blue, and the fraction numbers are shown along the X-axis in red.

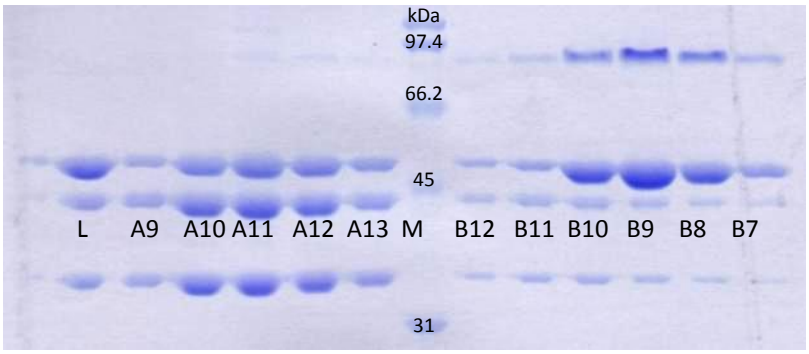



**Figure 7.13 SDS-PAGE of the gel filtration fractions.** The lanes are labelled with the fraction number corresponding to the trace shown in Figure 7.12. The standard protein markers are shown with their molecule weights (kDa) in the lane labelled M.

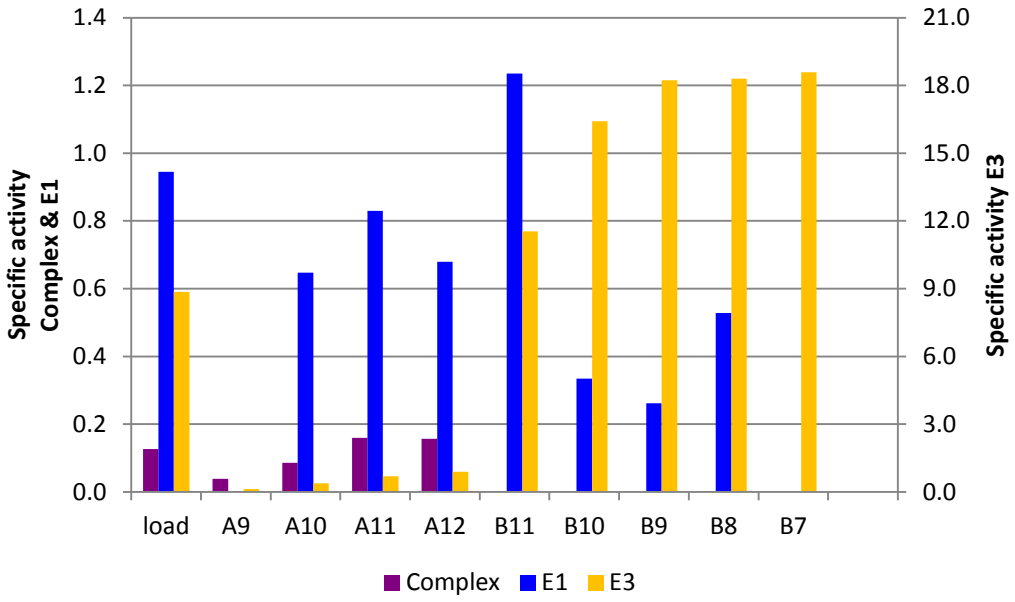



**Figure 7.14 A bar chart illustrating the specific activities of the E1, E3, and OADHC in each gel filtration fraction.** The specific activities are shown as U/mg of total protein where 1Unit is equivalent to 1µmol of product per min.

The two peaks shown in Figure 7.12 correspond to protein assemblies with an approximate molecular weight of 920 kDa and 88 kDa. The E2-trunc alone eluted from the gel filtration after 11.2ml, the approximate molecular weight of this protein peak is 340 kDa (Figure 6.9). When the E2-trunc was mixed with E1 and E3 there is no longer a peak at 11.2ml; instead the E2-trunc eluted after 8.95ml, a peak that contained branched-chain OADHC activity. This indicates that all the E2 trimers have bound E1 and E3 enzymes producing a trimeric OADHC (Figure 7.14). The second peak, which eluted between 13 and 15.5ml, contained the unbound E1$\alpha_2\beta_2$ and E3$\alpha_2$. More of the E1 than the E3 enzyme remains bound to the E2, this is the same pattern as was seen for the 42-mer complex, again suggesting that the E3 binds less tightly to the E2 than the E1 (Chapter 3.3.5).

It is difficult to predict the stoichiometries of the E1:E2:E3 from the SDS-PAGE, as the E2 and E3 monomers have similar molecular weights; however, it can be concluded that under these conditions the E2-trunc binds more E1 than E3. This gel filtration result cannot be directly compared with the gel filtration of the 42-mer complex as the previous experiment was carried out in a different buffer due to the stability of the E2-trunc protein, and this may have influenced the binding of the E1 and E3 to the E2-PSBD.

The specific activity of the trimeric OADHC is higher after gel filtration (fractions 11 and 12) as the excess unbound and thus non-functional E1 and E3 has been removed. The specific activity seen in this experiment is higher than the values previously achieved; this may be due to the assembly conditions in that the complex components were at a much higher concentration than that used for the preceding experiments and therefore more functional OADHC is likely to form.

### 7.3.4 Thermostability and temperature optimum of the trimeric OADHC

#### *7.3.4.1 Temperature optimum of the trimeric OADHC*

The temperature optima ($T_{opt}$) of the trimeric OADHC (E2-trunc) and the 42-mer OADHC (E2) (Chapter 3.3.4) are both 65°C (Figure 7.15). Preventing the trimers from assembling into a 42-mer has no affect on the thermoactivity of the OADHC.

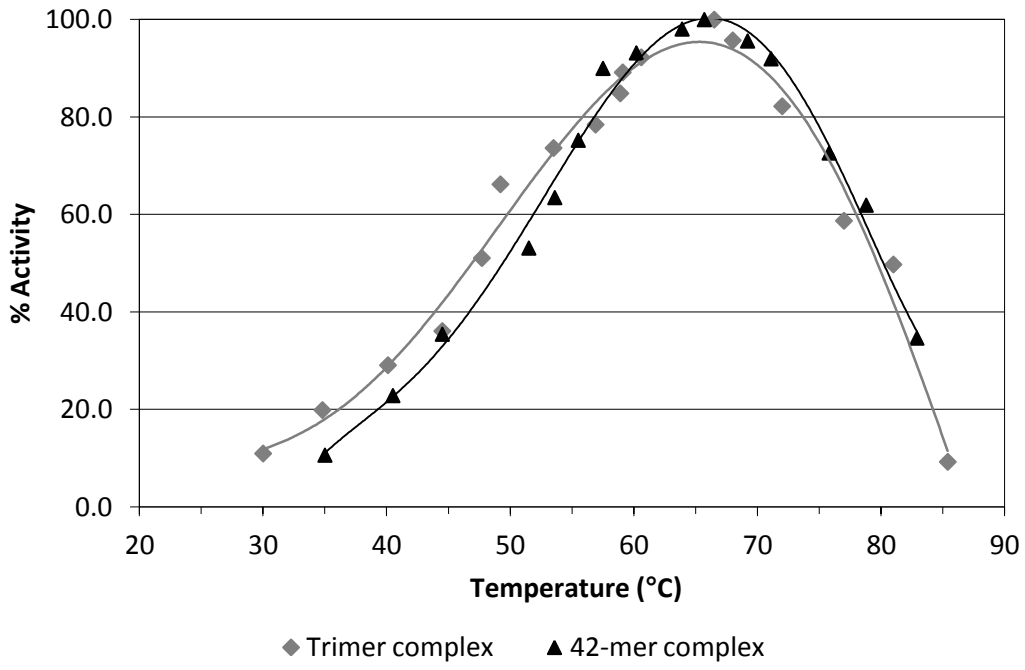



**Figure 7.15 Temperature optima of the trimeric and 42-mer OADHC.** The trimeric complex is composed of E1$\alpha_2\beta_2$:E2-trunc:E3$\alpha_2$ in stoichiometry of 1:1:0.5 and the 42-mer complex is composed of E1$\alpha_2\beta_2$:E2:E3$\alpha_2$ in a stoichiometry of 3:1:0.1.

### *7.3.4.2 Thermal inactivation of the trimeric OADHC*

The trimeric OADHC was assembled in a stoichiometry of 1:1:0.5 ($E1\alpha_2\beta_2$:E2-trunc:$E3\alpha_2$). There is very little difference between the thermostability of the non-truncated 42-mer E2 and the truncated trimeric E2 (Figure 7.4); however, the trimeric OADHC appears to be less thermostable than the 42-mer OADHC (Figure 7.17). The stoichiometry of the 42-mer complex was 3:1:0.5 (E1:E2:E3), whereas 1:1:0.5 was used for the trimeric complex; the affect that the stoichiometry of the OADHC has on thermostability is not known, and therefore the difference in stability between the two complexes may be due to the stoichiometry. The decrease in stability of the trimeric complex compared to the 42-mer may also be influenced by a change in interactions between the E2-trunc and the E1/E3 enzymes, or the binding of these enzymes to the E2 trimer may affect the strength of the interaction between the E2-trunc trimers. Any changes to the E2 stability could be investigated by repeating the E2-trunc thermal inactivation assays in the presence and absence of the E1 and/or E3 enzymes.

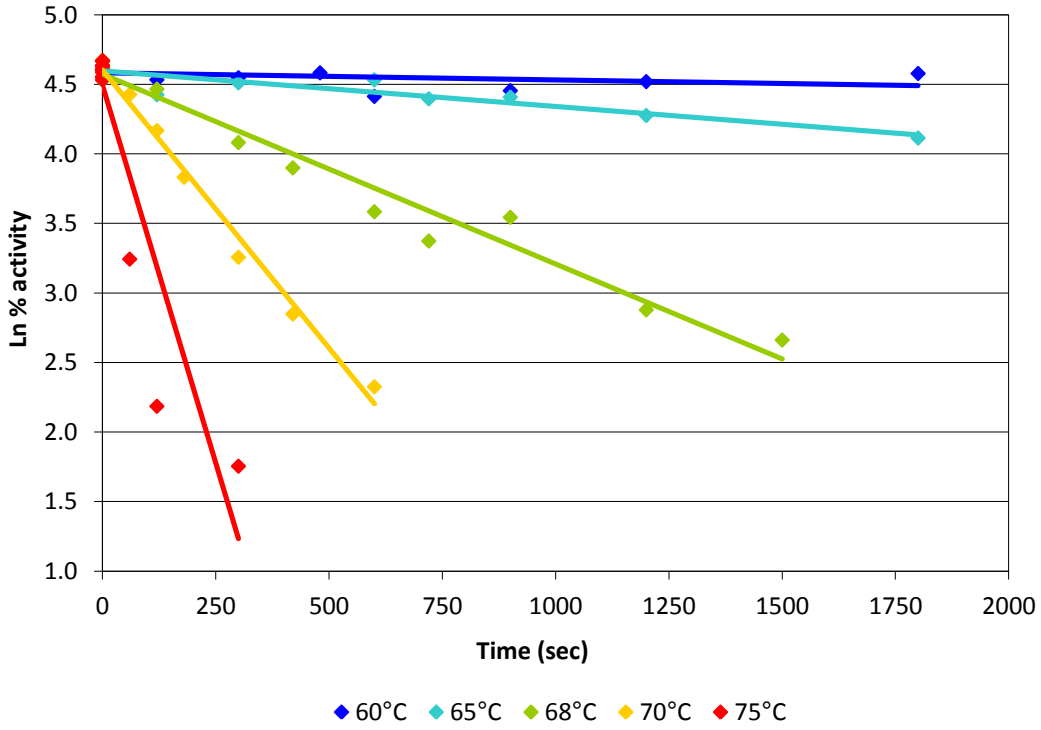



**Figure 7.16 Thermal inactivation profile of the trimeric OADHC.** The non-heated samples at time zero were taken to be 100% active. The ($E1\alpha_2\beta_2$:E2-trunc:$E3\alpha_2$) were mixed in a ratio of 1:1:0.5 and diluted 1:1 with 50mM potassium phosphate, pH 7.0, 0.2mM TPP, and 1mM $MgCl_2$.

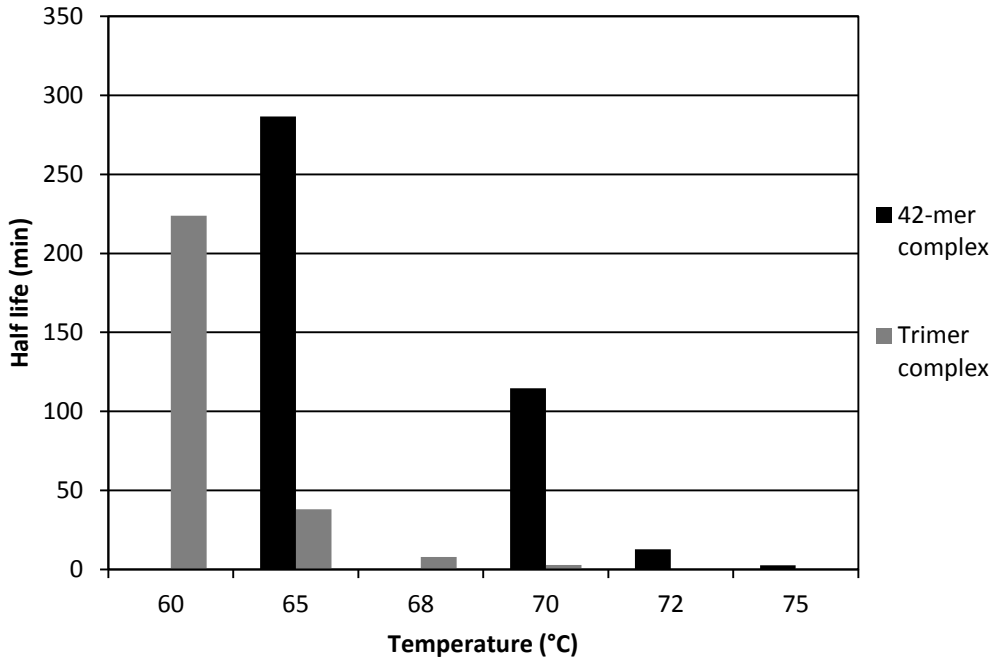



**Figure 7.17 Bar chart comparing the half-lives of the trimeric and wild type 42mer OADHCs.** The trimeric and 42-mer complexes were formed by mixing the following stoichiometries: trimeric complex ($E1\alpha_2\beta_2$:E2-trunc:$E3\alpha_2$) 1:1:0.5, and the 42-mer complex ($E1\alpha_2\beta_2$:E2:E3α2) 3:1:0.5.

## 7.4 Conclusion

The removal of the C-terminal anchor residue of the E2 protein does not alter the trimer structure, only its ability to form inter-trimer interactions. The trimeric E2-trunc and E2cat-trunc are capable of catalysing the dihydrolipoamide-dependent transacylase reaction, albeit at a reduced rate compared to the 42-mer E2. The trimeric E2 is also capable of binding E1 and E3, producing a functional trimeric OADHC; optimal activity was achieved when mixing the E1:E2-trunc:E3 enzymes in a stoichiometry of 1:1:1.

It would appear that it is not essential for the *Tp. acidophilum* branched-chain OADHC to assemble into a large 42-mer, but that assembly does increase the $V_{max}$ of the complex. The *Tp. acidophilum* trimeric branched-chain OADHC is currently the only example of a functional trimeric OADHC. Removal of the final C-terminal five residues (LLLDL) from the porcine E2, prevented 24-mer assembly but also abolished all E2 activity (Koike et al., 2000); although a trimeric *G. stearothermophilus* E2 has been reported (Peng et al., 2012), the effects on E2 and OADHC activity have not, to date, been published.

Direct comparisons between the trimeric and 42-mer OADHCs have been limited in this chapter due to variations in the experimental conditions used. Preferably all experiments would be performed in identical conditions utilising the same E1 and E3 preparations, so that any changes to the activity and stability are due only to the 42-mer and trimer assemblies rather than other factors.

# Chapter 8: The role of the 2-oxoacid dehydrogenase multienzyme complexes in *Tp. acidophilum* and other Archaeal OADHCs

## 8.1 Introduction

### 8.1.1 The role of the *Thermoplasma acidophilum* branched-chain OADHC

It was previously thought, that all archaea utilise ferredoxin oxidoreductases to catalyse the conversion of the various 2-oxoacids to their corresponding acyl-CoAs; however, genome sequencing has identified the putative genes encoding the proteins required to produce a fully-assembled and functional OADHC in numerous species of aerobic archaea. The *Tp. acidophilum* OADHC genes have been recombinantly expressed and the component enzymes assembled, resulting in a functional branched-chain OADHC (Heath et al., 2007). Dihydrolipoamide dehydrogenase (E3) activity has been detected in *Tp. acidophilum* cell lysates (Smith et al., 1987), but OADHC activity has not been detected. One of the aims of this chapter was to investigate a possible role of the *Tp. acidophilum* branched-chain OADHC within central metabolism*;* a genome search was performed in order identify genes encoding putative enzymes involved in the metabolism of branched-chain amino acids, details of this pathway can be found in Chapter 1.2.3 (Figure 1.4). The *Tp. acidophilum* OADHC has activity with both the branched-chain 2-oxoacids and pyruvate; as yet, it is unclear which substrate is the physiological substrate and under what conditions the OADHC replaces or complements the corresponding ferredoxin oxidoreductase enzyme.

### 8.1.2 The putative *Sulfolobus solfataricus* OADHC

The second aim of this chapter was to recombinantly express a hyper-thermostable OADHC from *S. solfataricus*, a hyper-thermophilic Crenarchaeota that grows at temperatures up to 85°C and between pH 2-3 (Brock et al., 1972). The E1 enzyme from this putative OADHC has been recombinantly expressed and been shown to possess a substrate specificity for acetoin, suggesting that *S. solfataricus* may possess an acetoin OADHC (Payne et al., 2010).

### 8.1.3 A putative *Picrophilus torridus* OADHC

The *Tp. acidophilum* E2 forms a 42-mer structure comprising both square and pentagonal faces, unlike the other previously described 24-and 60-mer structures. The final aim of this chapter was to identify, clone and express the OADHC genes from *P. torridus*; this organism grows optimally at 60°C and at pH 0.7, and along with *Tp. acidophilum* is of the order Thermoplasmatales (Futterer et al., 2004). The substrate specificity and E2 structure of the *P. torridus* putative OADHC may provide an insight into the evolutionary origins of the OADHC in Thermoplasmatales and the novel 42-mer E2 structure seen in *Tp. acidophilum*.

## 8.2 Methods

### 8.2.1 Identification of archaeal OADHCs and other genes

The Protein Basic Local Alignment Search Tool (PBLAST) programme was initially used to identify OADHC genes in other archaeal species, using the *Tp. acidophilum* OADHC genes as a reference sequence. Once one of the complex genes, usually E2, had been identified the Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Ogata et al., 1999) was used to identify and position the other OADHC component genes. KEGG was also used to identify other OADHC genes distal from any putative complex operon.

The putative enzymes involved in branched-chain amino acid metabolism were identified based on gene annotations and PBLAST searches using the *Bos tauraus* and *Pseudomonas putida* GB1 protein sequences as a reference. Genes encoding putative *S. solfataricus* enzymes involved in acetoin anabolism were identified using the sequences from *Klebsiella terrigena* (Blomqvist et al., 1993).

### 8.2.2 Cloning of the OADHC genes for heterologous expression in *E. coli*

The cloning of the OADHC genes from *S. solfataricus* into various pET vectors along with the protein expression trials were performed by Dr. K. Payne (University of Bath). The *P. torridus* E1, E2 and E3 genes were cloned from *P. torridus* DSM 9790 (DSMZ, Germany) genomic DNA, which was extracted by Nicholas Morant (GeneSys Ltd., Surrey); the PCR primers used to amplify the genes are shown in Table 8.1. The *P. torridus* OADHC genes were ligated into pGem-T Easy and pET28a using *Nhe* I/*Nco* I and *Xho* I, following the method described in Chapter 4.2.1.

| Primer name | Restriction Site | Primer Sequence 5'- 3' |
|---|---|---|
| E1α Forward | *Nhe*I | CGGCTAGCATGATTGAGGAAGATATTAGTAAAGAGG |
| E1α Reverse | *Xho*I | CCCTCGAGTCATTCTATCTCACCCCTTTCC |
| E1β Forward | *Nco*I | CCCATGGGCACCGAGATGAACATGGTAAAGG |
| E1β Reverse | *Xho*I | CCCTCGAGCTAATAATTTATAACCTTATTTATAGCATTCATAATTC |
| E2 Forward | *Nhe*I | AGCTAGCTAGCATGTACACGTTGAAGGTTCCACCAATAGG |
| E2 Reverse | *Xho*I | CCGCTCGAGTTAATCACCAATGTAAAGCATTGGC |
| E3 Forward | *NheI* | CGGCTAGCATGGATTACGATGTCATTATCCTTGG |
| E3 Reverse | *XhoI* | CCCTCGAGTTAATTTTCGAAGTCATCATGTATCCC |

**Table 8.1 The primers used to PCR amplify the *P. torridus* OADHC genes**. The sequence complementary to the gene is shown in black, the restriction site in red and any extra bases in blue. An additional G, shown in green, was added after the *Nco*I site and before the E1β gene to keep the gene in-frame with the ATG start codon found in the restriction site.

### 8.2.3 Expression of the OADHC proteins

The OADHC genes from *P. torridus* and E2 proteins from *S. solfataricus*, were expressed in *E. coli* Rosetta(DE3)pLysS following the method described in Chapter 2.3.1. Expression trials were conducted to attempt to increase expression of soluble *P. torridus* OADHC proteins using the following expression strains: *E. coli* BL21(DE3) grown in Overnight Express™ Instant TB medium (Novagen, Merck, U.K.) and ArcticExpress™ (Stratagene, Agilent Technologies, U.K.) grown in LB medium (Chapter 5.2.1.1).

### 8.2.4 Purification of the E2 dihydrolipoamide transacetylase

The *E. coli* cell pellets were resuspended in His-bind buffer and lysed by sonication as described in Chapter 2.3.2; the proteins were then purified by a heat precipitation at 70°C for 10 min, except those expressing the *P. torridus* E2 enzyme which was not heat treated, followed by nickel affinity chromatography (Chapter 2.3.2.2).

### 8.2.5 Characterisation of the E2 dihydrolipoamide transacetylase protein

#### *8.2.5.1 Dihydrolipoamide S-acetyltransferase spectrophotometric coupled assay*

The E2 activity assays were performed as described in Chapter 2.5.2.

#### *8.2.5.2 Determining the molecular weight of the assembled E2 protein*

The molecular weight of the assembled E2 protein was determined by gel filtration and dynamic light scattering (DLS); for details see Chapters 2.3.2.3 and 4.3.4.3, respectively.

#### *8.2.5.3 Thermostability of the E2 protein from Picrophilus torridus*

The thermostability of the *P. torridus* E2 was measured by heating aliquots of the protein and comparing the relative amounts of soluble, insoluble and aggregated protein by SDS-PAGE and DLS, following the methods described in Chapters 3.2.5, and 4.3.4.3, respectively.

### 8.2.6 Homologous expression of *S. solfataricus* putative E3 dihydrolipoamide dehydrogenase

The *S. solfataricus* E3 was expressed recombinantly following the method developed by Albers et al. (2006).

#### *8.2.6.1 Creation of the E3 expression plasmid for homologous expression*

The N-terminal his-tagged putative E3 (SSO1524) gene (Payne et al., 2010) was cloned by Dr. Karl Payne into the modified viral vector pMJ0503 (kindly provided by Dr. Sonja-Verena Albers, Max-Planck-Institute for Terrestrial Microbiology, Marburg, Germany) following the method developed by Albers et al. (2006). The E3 gene was cloned via pET28a via pGem-T. The E3 gene was then excised from pET28a along with the N-terminal his-tag and ligated into the pMZ1 vector; this vector added the arabinose promoter (AraS) region upstream of the his-tagged E3 gene. The E3 gene and AraS sequence were then excised and ligated into pMJ0503 to create the final homologous expression vector (Figure 8.1).

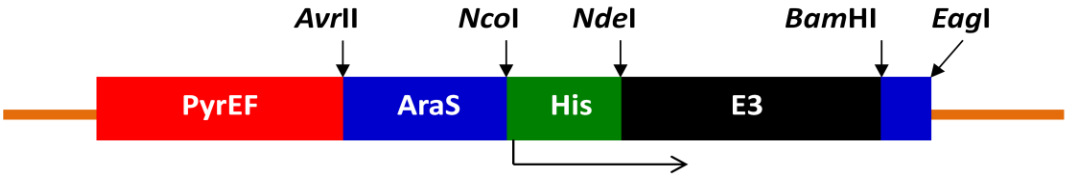



**Figure 8.1 *S. solfataricus* pMJ0503-E3 construct**. The pMJ0503 viral vector is shown in orange and contains the *S. solfataricus* E3 gene (black), the N-terminal his-tag (His) originally from pET28a (green), the arabinose promoter (AraS) (blue), and the marker gene PyrEF (red). The restriction enzymes used to clone each part are shown above and the arrow indicates the direction of transcription of the his-tagged E3 gene.

The pMJ0503-E3 viral vector was transformed into the *S. solfataricus* uracil auxotroph strain (ΔpyrEF/lacS double mutant) by Dr. Sonja-Verena Albers. The transformants were then grown under selective conditions using basal salt medium without uracil (Brock et al., 1972), containing (per litre): 0.28g $KH_2PO_4$, 1.3g $(NH_4)_2SO_4$, 0.25g $MgSO_4.7H_2O$, 0.25g $CaCl_2.7H_2O$ and adjusted to pH3.5. The medium was supplemented with 1% trace elements solution (20.0g $FeCl_2.6H_2O$, 4.5mg $Na_2B_4O_7.10H_2O$, 1.8g $MnCl_2.4H_2O$, 0.22mg $ZnSO_4.7H_2O$, 0.05mg $CuCl_2$, 0.03mg $Na_2MoO_4.4H_2O$, 0.03mg $VOSO_4.2H_2O$, 0.01mg $CoSO_4.7H_2O$), 0.05% (w/v) Yeast extract and 0.2 % D-arabinose (Albers et al., 2006). Selection of transformants was by growing cells on medium lacking uracil, the pMJ0503 plasmid contains the pyrEF gene which rescues the uracil auxoptroph mutation.

#### *8.2.6.2 Homologous expression of the S. solfataricus E3*

The transformed *S. solfataricus* cultures were grown in increasing volumes of medium until two 500ml cultures were inoculated and grown at 75°C; gene expression was induced by presence of D-arabinose in the medium during this time. The cultures were harvested after 2-4 days by centrifugation at 13400 x g, for 10 min at 4°C. The cells were resuspended in His-bind buffer and lysed by sonication; the his-tagged E3 was then purified from cell lysates following the method described in Chapter 2.3.2, except in this case, the heat purification step was omitted.

#### *8.2.6.2 E3 dihydrolipoamide dehydrogenase assay*

E3 dihydrolipoamide dehydrogenase activity was measured following the method described in Chapter 2.5.3, except the assay temperature was increased to 70°C.

## 8.3 Results and discussion

### 8.3.1 *Tp. acidophilum* branched-chain amino acid metabolism

The protein sequences from *B. taurus* and *P. putida* GB-1 were used to search the *Tp. acidophilum* genome for putative genes encoding enzymes involved in the metabolism of the branched-chain amino acids, Leu, Ile and Val into acetyl-CoA, acetoacetate, or succinyl-CoA (Chapter 1.2.2); these organisms were chosen as the branched-chain OADHC from each has been characterised (Kato et al., 2006, Ævarsson et al., 1999). Any hypothesis regarding the metabolism of the branched-chain amino acids in *Tp. acidophilum* is cautiously proposed, as the proteins have not been characterised and therefore catalytic activity and substrate specificity is not certain.

The branched-chain amino acids are initially converted into 2-oxoacids by a branched-chain aminotransferase (EC 2.6.1.42 and 2.6.1.6); the *Tp. acidophilum* genome does not contain any genes with homology to either the *B. taurus* or *P. putida* branched-chain aminotransferase enzymes. There are, however, nine genes in the *Tp. acidophilum* genome annotated as aminotransferases or aminotransferase-related proteins; the substrate specificity of these enzymes has not been determined, and it is therefore possible that one may possess branched-chain aminotransferase activity.

The *Tp. acidophilum* genes identified as possibly encoding enzymes involved in the metabolism of branched-chain acyl-CoAs are shown in Table 8.2. These results suggest that the *Tp. acidophilum* possesses several of the necessary enzymes but not all. From the data available, it is not possible to conclude if *Tp. acidophilum* utilises the branched-chain OADHC as part of the metabolism of Val, Leu and Ile, or if pyruvate is the OADHCs physiological substrate. There are two sets of ferredoxin oxidoreductase subunits genes annotated as having putative activity with pyruvate and 2-oxoglutarate, although interestingly there is not a set of genes annotated as a branched-chain FOR. The branched-chain 2-oxoacid degradation pathway can produce various molecules required for the biosynthesis of polyketides and branched chain fatty acids; however, *Tp. acidophilum* is unlikely to utilise these precursors as archaea do not synthesise polyketides and their cell membranes comprise isoprenoid lipids, rather than fatty acid lipids (Boucher et al., 2004).

Confirmation of the ability of *Tp. acidophilum* to metabolise branched-chain amino acids as a carbon source could be analysed by gene knock-outs and growth studies, as in the

case of *Haloferax volcanii,* in this case it has been suggested it utilises its branched-chain OADHC when grown on isoleucine but not valine or leucine (Sisignano et al., 2010).

| **Required enzyme** | ***Thermoplasma acidophilum* gene(s)** | |
|---|---|---|
| | **Gene number** | **Putative enzyme function** |
| **Isoleucine** | | |
| Acyl-CoA dehydrogenase (1.3.8.1) | Ta0230 | Isovaleryl-CoA dehydrogenase |
| | Ta0974 | acyl-CoA dehydrogenase |
| | Ta0295m | Acyl-CoA dehydrogenase |
| 3-Hydroxyacyl-CoA dehydrogenase (4.2.1.17) | Ta0121 | 3-hydroxybutyryl-CoA dehydratase-like protein |
| | Ta0291 | L-3-hydroxyacyl-CoA dehydrogenase precursor |
| | Ta0061 | 3-hydroxybutyryl-COA dehydratase |
| 3-Hydroxyacyl-CoA dehydrogenase (1.1.1.35) | - | - |
| Acyl-CoA thiolase (2.3.1.16) | Ta0296 | acetyl-CoA acetyltransferase |
| Propionyl-CoA carboxylase (6.4.1.3) | - | - |
| Methylmalonyl-CoA epimerase (5.1.99.1) | - | - |
| Methylmalonyl-CoA mutase (5.4.99.2) | Ta0462 | Methylmalonyl-CoA mutase, alpha subunit, N-terminus |
| **Valine** | | |
| Butyryl-CoA dehydrogenase (1.3.8.1) | Ta0230 | Isovaleryl-CoA dehydrogenase |
| | Ta0974 | acyl-CoA dehydrogenase |
| 3-Hydroxyacyl-CoA dehydrogenase (4.2.1.17) | Ta0121 | 3-hydroxybutyryl-CoA dehydratase-like protein |
| | Ta0291 | L-3-hydroxyacyl-CoA dehydrogenase precursor |
| 3-Hydroxyisobutyryl-CoA hydrolase (3.1.2.4) | - | - |
| 3-Hydroxyacyl-CoA dehydrogenase (1.1.1.31) | Ta0161 | Hypothetical protein |
| Methylmalonate-semialdehyde dehydrogenase (1.2.27) | Ta0809 | Aldehyde dehydrogenase |
| Propionyl-CoA carboxylase (See isoleucine) (6.4.1.3) | - | Valine pathway merges with isoleucine pathway at this step |

**Table 8.2 The enzymes required to metabolise the branched-chain acyl-CoAs produced by the branched-chain OADHC.** (Continued)

| Required enzyme | *Thermoplasma acidophilum* gene(s) | |
|---|---|---|
| | **Gene number** | **Putative enzyme function** |
| **Leucine** | | |
| Isovaleryl-CoA dehydrogenase (1.3.8.4) | Ta0230 | Isovaleryl-CoA dehydrogenase |
| | Ta0974 | acyl-CoA dehydrogenase |
| Methylcrotonyl-CoA carboxylase (6.4.1.4) | - | |
| γ-Carboxygeranoyl-CoA hydratase (4.2.1.18) | Ta0291 | L-3-hydroxyacyl-CoA dehydrogenase precursor |
| Hydroxymethylglutaryl-CoA lyase (4.1.3.4) | - | |

**Table 8.2 The enzymes required to metabolise the branched-chain acyl-CoAs produced by the branched-chain OADHC**. The required enzymes, along with their EC numbers, were identified from *B. taurus* or *P. putida* GB-1; the *Tp. acidophilum* gene number and putative enzyme function are shown in each case. The coloured *Tp. acidophilum* genes indicate where the same gene has been identified in more than one pathway. The valine pathway merges with the isoleucine pathway at the propionyl-CoA carboxylase step.

### 8.3.2 *S. solfataricus* acetoin dehydrogenase multienzyme complex

Please note that some results in this section include experiments completed by Dr. Karl Payne (University of Bath).

#### *8.3.2.1 Possible Acetoin Dehydrogenase multienzyme complex operon*

The genes required to produce a putative OADHC have been identified in *S. solfataricus*, Figure 8.2 shows the location of 4 of these genes The E2 gene contains a -1 frame-shift that results in a premature stop codon; this frame-shift is probably corrected in *S. solfataricus* during transcription by a process called programmed ribosomal frame-shift (Cobucci-Ponzano et al., 2006). The E2 gene used to express the recombinant E2 enzyme was mutated, by site directed mutagenesis, to remove this frame-shift, based on a sequence alignment with the *Sulfolobus islandicus* E2 gene which does not have the -1 mutation (experiments completed by Dr. Karl Payne).

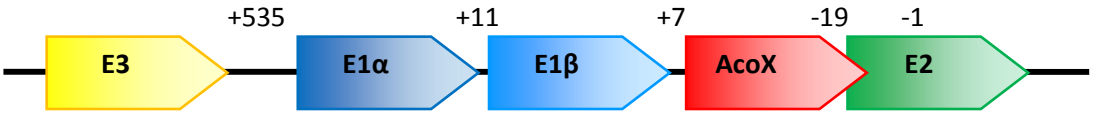



**Figure 8.2 OADHC genes in *Sulfolobus solfataricus* P2.** The numbers above the genes indicate the position of the gene relative to the end of the previous gene. The -1 above the middle of the E2 gene indicates that there is a -1 frame-shift mutation in the middle of this gene sequence.

There are also other sequences predicted to encode OADHC proteins located elsewhere in the genome. Another E1α and E1β sequence is found approximately 170kbp upstream, and there are three other E3 genes, with one located 410kbp upstream, and the other two located 30 and 950kbp downstream of this predicted operon.

#### *8.3.2.2 Expression and characterisation of the S. solfataricus OADHC proteins*

The OADHC genes have been cloned into various pET-vectors for recombinant expression in *E. coli*. The E1, E2 and E3 proteins were all recombinantly expressed, resulting in soluble E1 and E2 proteins, but insoluble E3 under all conditions tried. The E1 has been well characterised by Dr. Karl Payne as an acetoin $E1\alpha_2\beta_2$ decarboxylase (Payne et al., 2010).

### 8.3.3 Characterisation of the *S. solfataricus* E2

#### *8.3.3.1 Expression and purification*

The *S. solfataricus* E2 protein was expressed from pET28a in *E. coli* Rosetta(DE3)pLysS cells; the protein was then purified by heat precipitation of the *E. coli* proteins and nickel affinity chromatography, resulting in a protein that was approximately 90% pure (Figure 8.3).

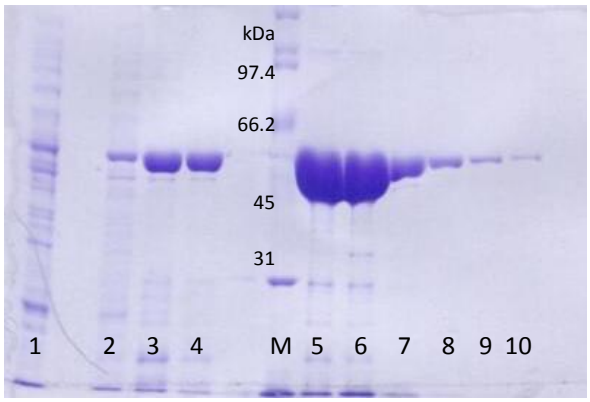



**Figure 8.3 SDS-PAGE of the nickel affinity chromatography purification of the *S. solfataricus* E2**. Lane 1: flow-through protein that did not bind to the column; lanes 2 to 4: the his-bind buffer, 5% and 10% his-elute washes; lanes 5 to 9: his-elute wash collected in 1ml fractions. The standard protein markers, with their molecular weights (kDa), are shown in the lane labelled M.

#### *8.3.3.2 E2 dihydrolipoamide acyltransferase enzyme assays*

The *S. solfataricus* E1 has been shown to utilise acetoin. In this case, unlike other E1 catalysed reactions, acetaldehyde is released rather than carbon dioxide; this leave an acetyl group attached to the lipoyl domain as would occur if pyruvate was the substrate. An acetoin specific OADHC would catalyse the following reaction (Payne et al., 2010):

$$\text{Acetoin} + \text{CoA} + \text{NAD}^+ \rightarrow \text{acetadehyde} + \text{acetyl-CoA} + \text{NADH} + \text{H}^+$$

The substrate specificity of the E1 enzyme indicates that the E2 will be able to catalyse the transfer of an acetyl group, so this E2 can be assayed using the phosphotransacetylase-E2 coupled assay.

| Organism | Specific activity (U/mg) | $k_{cat}$ ($min^{-1}$) |
|---|---|---|
| ***Tp. acidophilum* E2** | 1.1 ± 0.1 | 48.9 ±2.3 |
| ***S. solfataricus* E2** | 10.4 ± 0.6 | 678.7 ± 54.2 |

**Table 8.3 Specific activity and $k_{cat}$ values of the E2 dihydrolipoamide transacetylase from *Tp. acidophilum* and *S. solfataricus* at 55°C.** The specific activity is shown in U/mg of E2, where 1 Unit is described as the conversion of 1 µmol of acetyl-CoA to 1 µmol of acetyl-dihydrolipoamide per min. The $k_{cat}$ value is the turnover number per min per E2 monomer.

The specific activity and $k_{cat}$ values of the non-lipoylated E2 enzymes from *S. solfataricus* and *Tp. acidophilum,* at 55°C, are shown in Table 8.3. The *S. solfataricus* E2 has a higher specific activity and $k_{cat}$ than the *Tp. acidophilum* E2, even at 55°C a temperature that may be lower than *S. solfataricus* E2's optimum. It is not unexpected that the S*. solfataricus* E2 has a higher specific activity as the *Tp. acidophilum* is a branched-chain OADHC which only possesses low levels of activity with pyruvate.

#### *8.3.3.2 Molecular weight of the S. solfataricus E2*

The multimeric nature of the *S. solfataricus* E2 was investigated to determine if it assembled into a 24, 42 or 60-mer. A sample of the *S. solfataricus* E2 was purified by nickel affinity chromatography and analysed by size-exclusion gel filtration in 50mM Tris-HCl, pH 8.5, 200mM NaCl, and 100mM imidazole. The E2 protein eluted from the column in a peak at 13.09 ml (Figures 8.4 and 8.5), this equates to a molecular weight of approximately 150 kDa. The E2 monomers have a predicted molecular weight of 46.4 kDa, and therefore the gel filtration results suggest that the *S. solfataricus* E2 assembles into a structure comprising 3.3 subunits; this result must be viewed cautiously, as has been seen with the *Tp. acidophilum* E2, the E2's extended flexible linker regions can inaccuracies when determining the molecular weight of this protein by gel filtration.

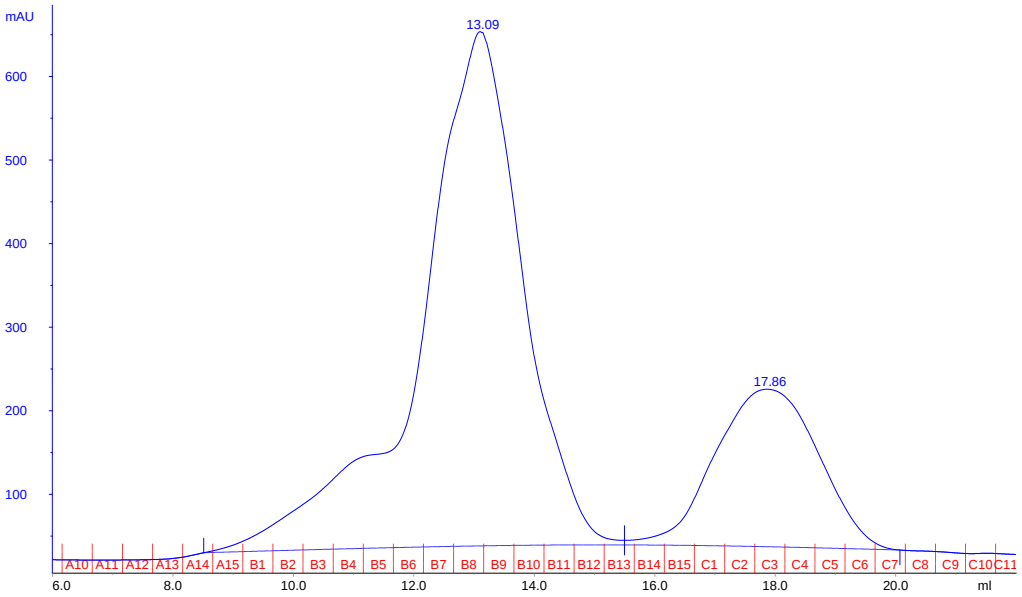



**Figure 8.4 Gel filtration trace of the *S. solfataricus* E2 protein**. The $A_{280}$ readings are shown in blue, and the 0.5ml fractions shown in red along the X-axis; no protein peaks were collected prior to those shown. The column was equilibrated and run in 50mM Tris-HCl, pH8.5, 200mM NaCl, and 100mM imidazole.

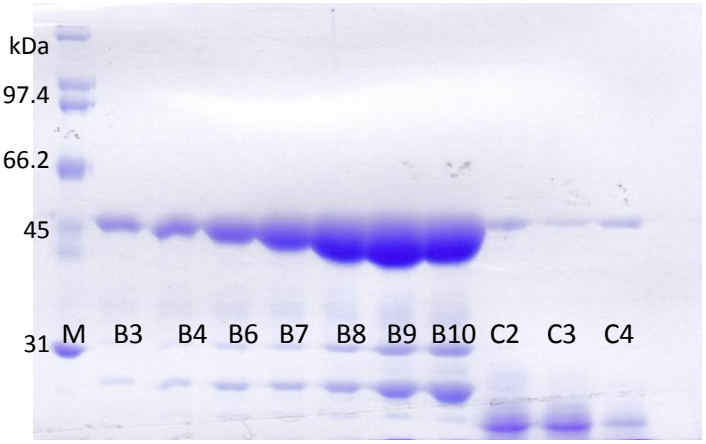



**Figure 8.5 SDS-PAGE of gel filtrations fractions.** The lanes are labelled with the fraction number corresponding to Figure 8.4 The lane labelled M contains standard protein markers along with their molecular weights (kDa).

Dynamic light scattering was also used to determine the molecular weight of the *S. solfataricus* E2 protein. The E2 protein in 50mM Tris-HCl, pH 8.5, 200mM NaCl, and 100mM imidazole, was monodispersed, with a hydrodynamic radius of 5.84 nm; assuming that the E2 is a globular protein this size equates to a molecular weight of 209 kDa or 4.5 monomers.

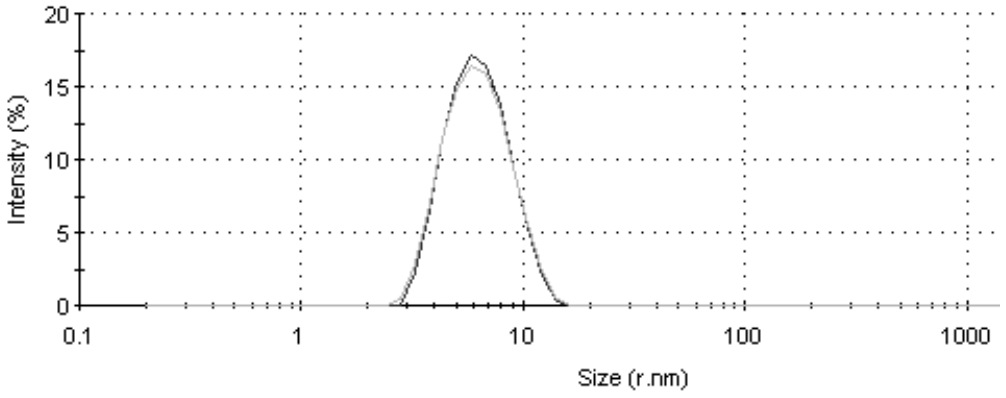



**Figure 8.6 DLS size distribution peak at 25°C.** The hydrodynamic radius of the protein sample is shown verses the signal intensity; duplicate measurements were taken and are shown in green and red.

The molecular weight of the E2 protein determined by gel filtration and DLS equate to a structure comprising 3.3 and 4.5 monomers, respectively. The variation between the two results means that the multimeric nature of the *S. solfataricus* cannot definitely be established; however, both results imply that the *S. solfataricus* E2 is not assembling into a large structure (24, 42 or 60-mer).

### 8.3.4. Homologous expression of the *S. solfataricus* E3

All attempts to recombinantly express the E3 gene, located 535bp upstream of the E1α gene, resulted in insoluble protein; all attempts to refold the insoluble E3 protein were unsuccessful. As the E1 and E2 proteins were soluble and active, homologous expression of the his-tagged *S. solfataricus* E3 was attempted with assistance from Dr. Karl Payne (University of Bath) and Dr. Sonja-Verena Albers (Max Planck Institute for Terrestrial Microbiology, Marburg, Germany).

### *8.3.4.1 Cloning of the S. solfataricus E3 into pMJ0503*

The creation of the pMJ0503-E3 vector was completed by Dr. Karl Payne and the transformation of this plasmid into the *S. solfataricus* M16 homologous expression strain (Albers et al., 2006) was completed by Dr. Sonja-Verena Albers.

### *8.3.4.2 Homologous expression of the S. solfataricus E3*

The homologous expression of the *S. solfataricus* E3 gene was induced with 0.2% (w/v) D-arabinose. The his-tagged E3 was purified from cell lysates by nickel-affinity chromatography (Figure 8.7); a protein with the expected molecular weight of a his-tagged E3 was purified, and this putative E3 protein was soluble but very dilute (approximately 0.01mg/ml).

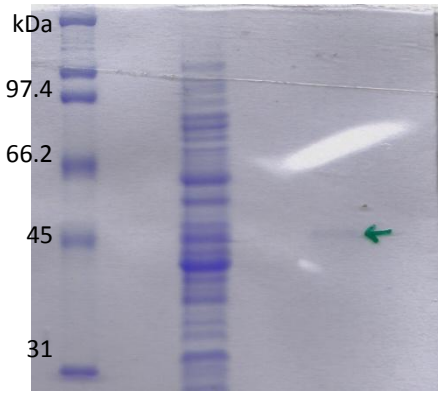



**Figure 8.7 SDS-PAGE nickel-affinity chromatography purification of the homologously expressed *S. solfataricus* E3.** Lane 1 contains the standard proteins along with their molecular weights (kDa); lane 2 contains the *S. solfataricus* cell lysate, and lane 3 contains protein the eluted from the nickel-affinity column with his-elute buffer. The band on lane 3 is very faint so its position has been highlighted with the adjacent arrow.

### *8.3.4.3. E3 dihydrolipoamide dehydrogenase*

The putative E3 protein was assayed for dihydrolipoamide dehydrogenase activity with $NAD^+$ and $NADP^+$, but none was detected. The protein was also incubated at 60°C and 70°C with 0.05mg/ml FAD and FMN, but this also did not result in any detectable activity. Unsurprisingly, no OADHC activity was detected when the E3 was mixed with heterologously expressed E1 and lipoylated E2. The lack of soluble and functional E3 may be due to the correct expression condition not yet being identified; however, it should also be noted that unlike several other species aerobic archaea dihydrolipoamide dehydrogenase activity has never been detected in *S. solfataricus* cell lysates

#### *8.3.4.4 The OADHC from S. solfataricus*

The *S. solfataricus* E1 substrate specificity and E2 activity implies that this complex catalyses the following reaction:

$$\text{Acetoin} + \text{CoA} + \text{NAD}^+ \rightarrow \text{Acetaldehyde} + \text{Acetyl-CoA} + \text{NADH+H}^+$$

*S. solfataricus* also possess an acoX gene (SSO1527) in between the E1β and E2 genes, the protein AcoX is predicted to be involved in the catabolism of acetoin (Priefert et al., 1991). These results possibly indicate that *S. solfataricus* may utilise acetoin as a carbon source; however, the source of this acetoin is not yet known. The organism may convert pyruvate to acetoin, as a method of storing carbon and preventing over-acidification of the cytoplasm; the enzymatic steps required to convert pyruvate to acetoin are shown in Figure 8.8. A gene annotated as encoding an acetolactate synthase (ALS) is present the *S. solfataricus* genome; however, no putative α-acetolactate decarboxylase (ADLH) has been identified (Payne et al., 2010). The lack of this gene does not discount the possibility of this pathway existing in *S. solfataricus,* as this reaction can occur spontaneously at high temperatures (Kobayashi et al., 2005). Acetoin production has not been detected in *S. solfataricus*, so the α-acetolactate may instead be diverted into the production of branched-chain amino acids rather than acetoin; if this is the case, then exogenous acetoin maybe the OADHC's physiological substrate, allowing the organism to utilise scavenged acetoin as an additional carbon energy source. As no E3 or OADHC activity has been detected in cell lysate, it is not known if or under what conditions *S. solfataricus* utilises this OADHC.

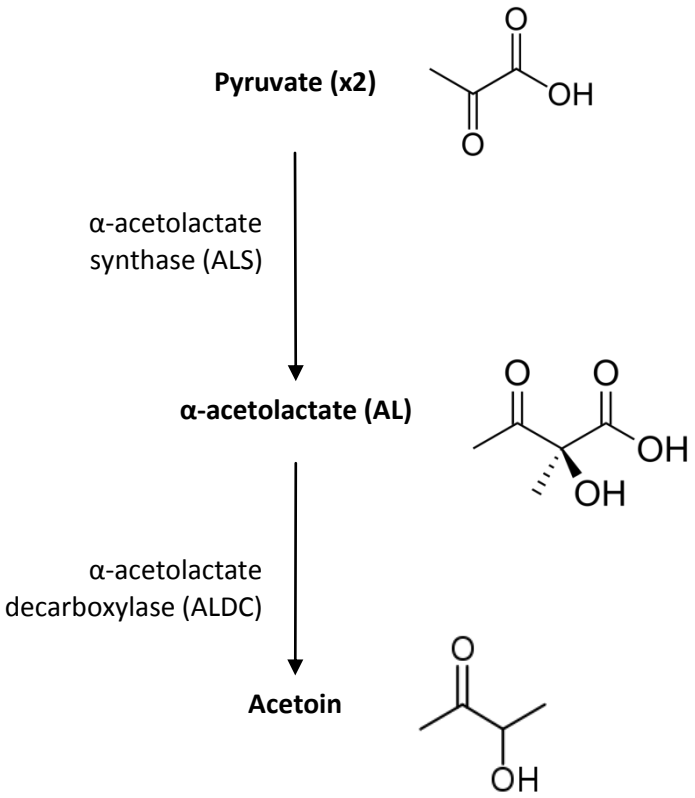



**Figure 8.8 Enzymatic conversion of pyruvate to acetoin**. Two pyruvate molecules can be converted via α-acetolactate into one acetoin molecule. α-acetolactate is an unstable molecule so the conversion to acetoin can occur non-enzymatically.

### 8.3.5 Cloning of the multienzyme complex genes from *P. torridus*

Please note that some results included are from experiments completed by Isabel Diez-Sevilla (Undergraduate student, University of Bath).

The OADHC genes have been identified in the genome of *P. torridus;* this organism grows in conditions similar to those inhabited by *Tp. acidophilum* (Darland et al., 1970, Futterer et al., 2004). The OADHC proteins from this organism may provide information regarding the evolutionary origins of the substrate specificity and structure of the *Tp. acidophilum* branched-chain OADHC.

#### *8.3.5.1 Identification of the OADHC genes in P. torridus*

The OADHC genes from *P. torridus* are located adjacent to each other, (Figure 8.9), in a similar arrangement to the *Tp. acidophilum* operon. The putative *P. torridus* operon is flanked by the lipoic acid synthase (LipA) and protein N-octanoyltransferase (LipB) genes; these genes are not present in *Tp. acidophilum,* which utilises a lipoate protein ligase A (LplA) system instead (Posner et al., 2009).

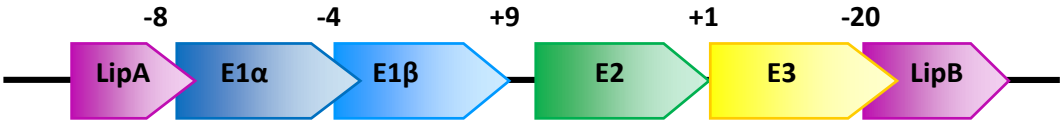



**Figure 8.9 OADHC genes in *P. torridus.*** The numbers above the genes indicate the position of the gene relative to the end of the previous gene. LipA is a predicted lipoic acid synthase and LipB is a protein N-octanoyltransferase.

#### *8.3.5.2 Cloning of OADHC genes*

The E1, E2 and E3 genes were successfully PCR-amplified from genomic DNA and ligated into pET28a (Figure 8.10). There is a 4bp overlap between the E1α and E1β genes, and analysis of the sequence upstream of the E1β start codon revealed a possible ribosome binding site and TATA box within the E1α gene. The E1α-E1β tandem (Figure8.10a) was cloned into pET28a, resulting in the co-expression of the his-tagged E1α and non-tagged E1β; the E1β gene was also cloned on its own into pET19b.

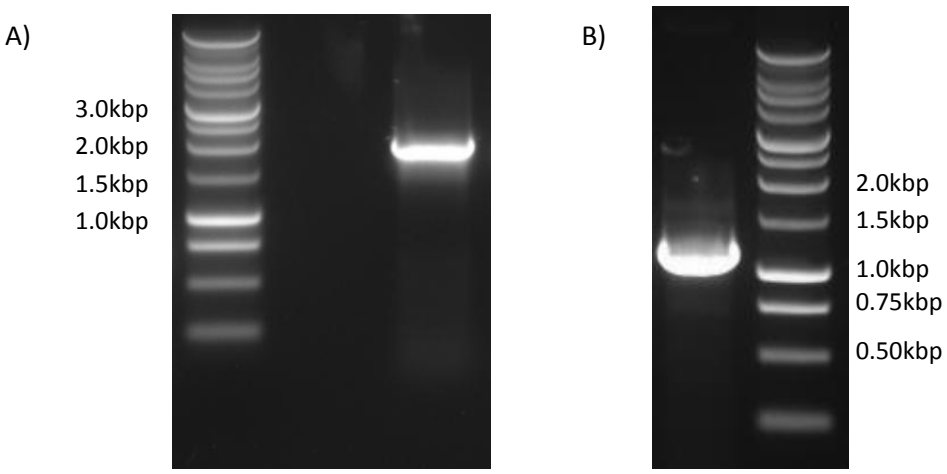



**Figure 8.10 Agarose gel of A) the E1αβ tandem and B) the E2 PCR products**. The DNA marker ladder on both gels is a 1Kb ladder (Promega, U.K.) The E1αβ has an expected size of 2.0kbp and the E2 has an expected size of 1.2kbp

### 8.3.6 Expression of *P. torridus* complex proteins

#### *8.3.6.1 Expression E1 and E3*

Expression of the E1α and E1β genes from the tandem E1αβ plasmid construct resulted in soluble E1α protein but insoluble E1β (Figure 8.11). Co-transformation of the pET28a-E1αβ and pET19b-E1β resulted in soluble E1α, and again insoluble E1β. Expression trials altering the expression strain and medium were conducted by Isabel Diez-Sevilla but did not result in the expression of any soluble E1β alone or in the presence of E1α. All attempts to express *P. torridus* E3 protein from pET28 using a range of expression strains and media resulted in no expression of the E3 protein in either a soluble or insoluble form.

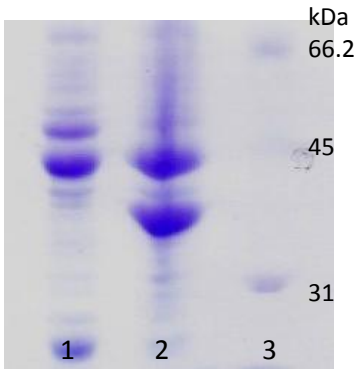



**Figure 8.11 SDS-PAGE of the soluble and insoluble fractions from a cell lysis containing the E1α and E1β proteins expressed from the E1αβ tandem**. Lane 1 contains the soluble cell lysate proteins, lane 2 contains the insoluble cell lysate proteins, and lane 3 contains the standard protein markers labelled with their molecule weights (kDa). The expected molecular weight of the E1α and E1β are 40.4 kDa and 35.3 kDa respectively.

### *8.3.6.2 Expression E2*

The E2 protein with an N-terminal his-tag was successfully expressed in *E. coli*. The protein was purified by nickel affinity chromatography, resulting in a final sample that was approximately 85% pure (Figure 8.12).

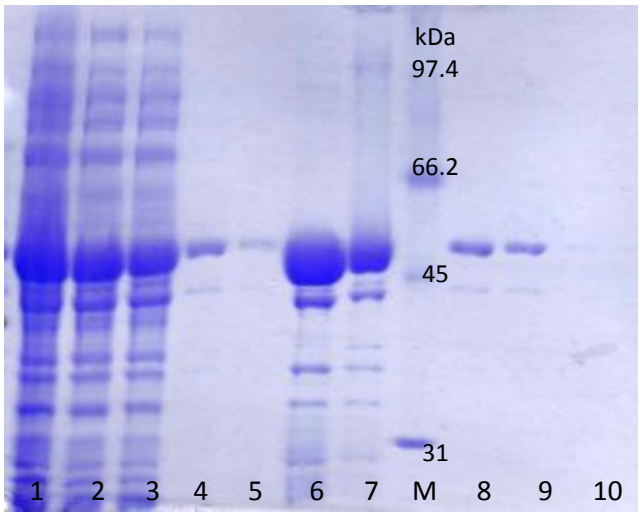



**Figure 8.122 SDS-PAGE of the purification of the E2 protein.** Lane 1: cell lysate loaded onto the column; lane 2: flow through; lanes 3: his-bind buffer wash; lanes 4 and 5: 5 and 10% his-elute washes; lanes 6 – 10: 1ml fractions of 100% his-elute buffer. The lane labelled M contains the standard protein markers, with their molecular weights (kDa). The his-tagged E2 protein has an expected molecular weight of 45.7 kDa.

### 8.3.7 Characterisation of *P. torridus* E2

#### *8.3.7.1 Sequence alignment of the P. torridus and Tp. acidophilum E2*

The *P. torridus* and *Tp. acidophilum* E2 protein sequences align with 43% identity, 62% similarity and 9% gaps. The key residues involved in assembly and catalytic activity have been highlighted in Figure 8.13. The flexible linker region that links the peripheral subunit binding domain (PSBD) and the catalytic domain appears to be significantly reduced in the *P. torridus* E2 protein; how the condensation of the PSDB and catalytic domains affects E1 and E3 binding along with OADHC activity has not yet been investigated.

**Figure 8.13 Sequence alignment of the E2 proteins from *P. torridus* and *Tp. acidophilum*.** Residues and sequences of known importance have been highlighted. The yellow lysine (K) located in the lipoyl domain is the site of lipoylation. The flexible linker regions of the *Tp. acidophilum* E2 have been highlighted in green (Lipoyl domain - PSBD) and cyan (PSBD - catalytic domain). The residues highlighted in red are key residues required for acyltransferase activity. The Ile highlighted in purple is the *Tp. acidophilum* anchor residue and the corresponding residue in the *P. torridus* sequence.

### *8.3.7.2 P. torridus E2 dihydrolipoamide transacetylase activity*

The *P. torridus* E2 was assayed using the coupled E2-PTA assay. The enzyme had a specific activity and $k_{cat}$ that were approximately 1.6-fold higher than the *Tp. acidophilum* E2, but the values were significantly lower those of the *S. solfataricus* E2 (Table 8.4). Comparing the $k_{cat}$ values of the *P. torridus* E2 with *Tp. acidophilum* and *S. solfataricus* values, it could be hypothesised that *P. torridus* OADHC may be specific for a substrate other than pyruvate; that is, as the rate of acetyl transfer is more similar to the *Tp. acidophilum* E2, than the *S. solfataricus* E2. The *P. torridus* OADHC may be specific for either the branched-chain 2-oxoacids, or 2-oxoglutarate. However, it is not possible to confirm this as the recombinant E1β is insoluble.

| Organism | Specific activity (U/mg) | $k_{cat}$ ($min^{-1}$) |
|---|---|---|
| ***P. torridus*** | 1.7 ± 0.1 | 76.4 ± 4.7 |
| ***Tp. acidophilum*** | 1.1 ± 0.1 | 48.9 ±2.3 |
| ***S. solfataricus*** | 10.4 ± 0.6 | 678.7 ± 38.3 |

**Table 8.4 Specific activity and $k_{cat}$ values of the E2 dihydrolipoamide transacetylase from *Picrophilus torridus*, *Thermoplasma acidophilum* and *Sulfolobus solfataricus*.** The specific activity is shown in U/mg of E2, where 1 Unit is described as the conversion of 1 µmol of acetyl-CoA to 1 µmol of acetyl-dihydrolipoamide per min. The $k_{cat}$ value is the turnover number per E2 monomer.

### *8.3.7.4 Size determination of the P. torridus E2*

A purified sample of the *P. torridus* E2 was loaded onto a gel filtration column equilibrated with 20mM Tris-HCl, pH8.5, 300mM NaCl, and 50mM imidazole. The E2 protein eluted from the column in the void volume of the column, from this, it can be concluded that the *P. torridus* E2 has assembled into a large structure but a precise molecular weight cannot be determined by this method (data not shown).

Dynamic light scattering was used to determine the hydrodynamic radius of the protein and therefore predict the molecular weight of the assembled structure. The hydrodynamic radius of the *P. torridus* E2 at 25°C in 20mM Tris-HCl, pH 8.5, 300mM NaCl, and 50mM imidazole, was 15.7 nm (Figure 8.14); assuming that the E2 forms a globular structure, this equates to a molecular weight of 2.1 MDa (the expected molecular weights of a 24, 42 and 60-mer are 1.1 MDa, 1.9 MDa and 2.7 MDa, respectively). The molecular weight of the E2 protein calculated from the hydrodynamic radius is likely to be overestimated, as the algorithm that converts the radius into a $M_w$ assumes that the proteins are both globular and solid. The flexible arms of the E2 extend out from the

catalytic core, and this increases the hydrodynamic radius to a size above that which would be expected of a globular compact protein of the same molecular weight.

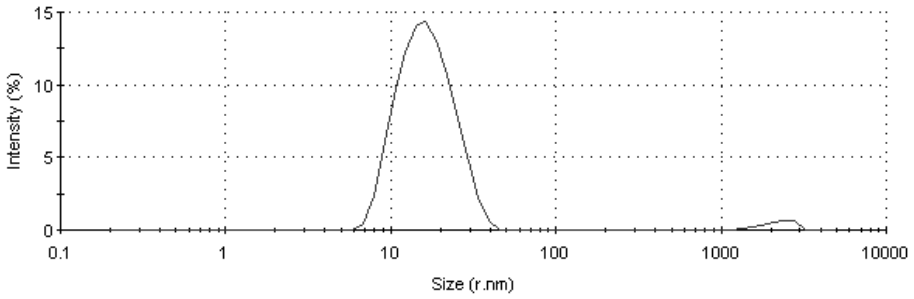



**Figure 8.14 DLS intensity graph of the *P. torridus* E2 protein at 25°C**

The hydrodynamic radius of the *Tp. acidophilum* E2 was determined by DLS to be 19.3 nm, which is larger than the radius of the *P. torridus* E2. It is important to note that the flexible linker region between the *P. torridus* catalytic domain and PSBD is shorter than the equivalent *Tp. acidophilum* linker; shortening this region may reduce the distance that the flexible arms project outward and thus decrease the hydrodynamic radius of the protein. The DLS result suggests that the *P. torridus* E2 is smaller than the *Tp. acidophilum* E2, but the exact number of monomers cannot be established. To determine the structure of the *P. torridus*, the molecular weight would need to be accurately measured by analytical ultracentrifugation, or the structure ascertained by either SAXS or X-ray crystallography.

#### *8.3.7.5 Thermostability of the P. torridus E2*

The thermostability of the *P. torridus* E2 protein was studied by comparing the relative amounts of soluble and insoluble protein, after a 10 min incubation at a range of temperatures (Figure 8.15). The *P. torridus* E2 is stable for up to 10 min at 80°C, but above this temperature the protein begins to precipitate; the majority of the protein is insoluble after 10 min at 95°C. This result is similar to that seen for the *Tp. acidophilum* E2; this protein began to precipitate after 10 min at 80-85°C and had completely precipitated after 10 min at 95°C (Chapter 3.3.2.3).

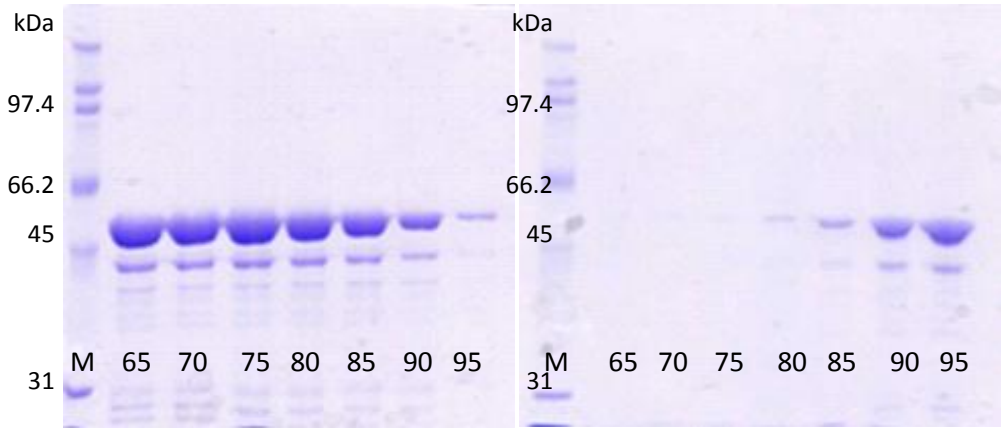



**Figure 8.15 SDS-PAGE of A) the soluble and B) the insoluble E2 protein after a 10 min incubation over a range of temperatures**. The lanes are labelled with the temperatures (°C) at which the samples were incubated for 10 min in 20mM Tris-HCl, pH 8.5, 300mM NaCl, and 50mM imidazole. The standard proteins are shown in the lanes labelled M along with their molecular weights (kDa).

The thermostability of the *P. torridus* E2 was also studied by DLS. The hydrodynamic radius of the E2 in 20mM Tris-HCl, pH 8.5, 300mM NaCl, and 50mM imidazole, was determined after one-minute incubations at a range of increasing temperatures between 50°C and 90°C. The average size begins to increase slightly between 75°C and 80°C, and then dramatically increases after 80°C, indicating that the protein had aggregated at this temperature (Figure 8.16). The thermostability of the *P. torridus* and the *Tp. acidophilum* E2 proteins measured by DLS and SDS-PAGE are similar, this is not unexpected as both organisms inhabit similar environments; the effect that temperature has on the activity of the *P. torridus* E2 has not yet been studied.

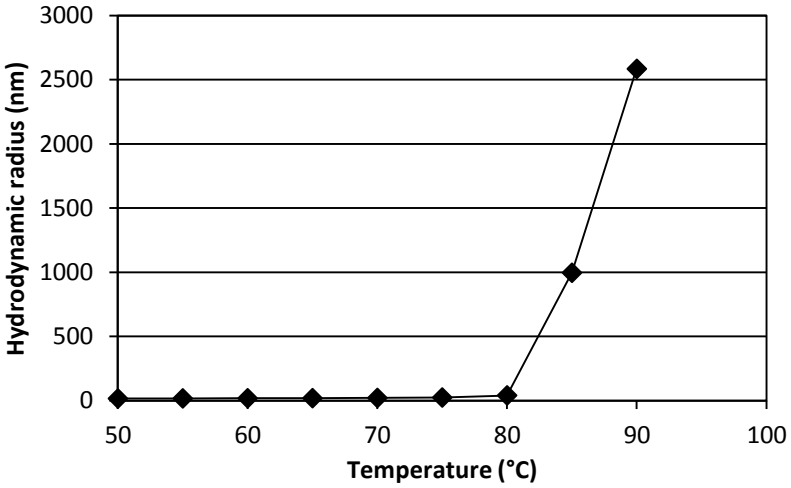



**Figure 8.16 The hydrodynamic diameter of the *P. torridus* E2 across an increasing temperature ramp.**

## 8.4 Conclusion

It is not possible, with the data currently available, to determine the metabolic role of the *Tp. acidophilum* branched-chain OADHC. Several putative genes have been identified as encoding enzymes involved in the metabolism of the branched-chain amino acids, but as these enzymes and the pathways have not be characterised, it is not possible to draw any definitive conclusions. The OADHC genes have been shown to be transcribed and translated in *P. aerophilum* and *Tp. acidophilum* (Cozen et al., 2009, Sun et al., 2011) but no OADHC activity has ever been detected in cell lysates. The physiological role these complexes play in archaeal metabolism is further complicated by the fact that archaea appear to utilise the ferredoxin oxidoreductase to catalyse the conversion of 2-oxoacids to their corresponding acyl-CoA. Nevertheless, as archaeal genomes have been streamlined during evolution (Csűrös and Miklós, 2009), and therefore these organisms do not maintain genes that are surplus to requirement, this suggests that the OADHC identified in *Tp. acidophilum* and the other aerobic archaea are likely to have a physiological function, otherwise functional E1, E2, E3 and the accessory OADHC specific protein lipoylation enzymes would not be so widely maintained.

The E1α, E1β, E2, and E3 proteins from *Tp. acidophilum* have been recombinantly expressed and assembled into an active OADHC; however, thus far it has not been possible to replicate this from another thermophilic or hyper-thermophilic archaeon. It seems most likely that this is due to the inability of the *E. coli* expression systems to express and fold these thermophilic proteins correctly, and although a *S. solfataricus* expression system has been developed (Albers et al., 2006) it is not yet as efficient as the *E. coli* system. Further understanding of the thermostability of these complexes may be achieved by studying the proteins from thermophilic bacteria, such as the *Geobacillus stearothermophilus* PDHC (60°C) or other uncharacterised complexes.

The nature of the *P. torridus* E2 assembly may provide some insights into the evolutionary origins of the novel *Tp. acidophilum* E2 42-mer structure. The *P. torridus* E2 has been recombinantly expressed in *E. coli* resulting in a functional enzyme that assembles into a multimeric structure, although the exact nature of this structure has not yet been determined. The production and characterisation of the *P. torridus* E2 catalytic domain would ideally be followed, if possible, by a high-resolution X-ray crystallographic structure.

# Chapter 9: Conclusions and future work

### 9.1 The recombinant OADHC from *Tp. acidophilum*

The *Thermoplasma acidophilum* OADHC enzymes have been recombinantly expressed in *E. coli*, resulting in a functional multienzyme complex (Heath, 2006); the assembled OADHC is both thermostable and thermoactive at temperatures above the organisms expected growth range. The temperature optimum of the OADHC appears to be limited by the E1 component, an enzyme that has previously been suggested to be the rate limiting step of the overall oxidative decarboxylation reaction (Pei et al., 2008). The assembled complex is less thermostable than the individual component enzymes; this insinuates that in this case it is the stability of the interactions of the E1/E3 with the E2, or between E2 trimers, that limits the thermostability of the complex. Assuming that the thermostability and thermoactivity of the recombinant OADHC reflects those of the *in vivo* enzyme, these data suggest that the *Tp. acidophilum* branched-chain OADHC would be stable and functional at the temperatures at which this organism grows; therefore, despite the fact that OADHC activity has not be detected *in vivo*, this complex could theoretically be utilised by the organism.

The assembly and the stability of the *Tp. acidophilum* OADHC have only briefly been studied in this project; the results suggest that the E1 and E3 do not bind strongly to the E2 core. The weak binding of the E3 to the E2 was not unexpected, as proteomic studies of the large complexes (>1 MDa) in *Tp. acidophilum* have only identified the E1 associated with the E2 protein, and not the E3 (Sun et al., 2007). One question that has not been answered is, *in vivo* could non-bound E3 enzymes re-oxidise the E2 lipoic acid moiety, and hence result in an E1 and E2 complex with E3 enzymes in close proximity? The binding affinities of the E1/E3 for the E2-PSBD could be calculated using surface plasma resonance, to determine if and how tightly the E3 associates with the E2. Weak interactions may be sufficient to achieve a functional complex within a cell lysate, but they may be easily broken during cell lysis, this may explain the lack of detectable OADHC activity when cell lysate have been assayed.

## 9.2 The structure of E2 core of the archaeal OADHCs

### *9.2.1 The novel 42-mer structure*

The structure of the *Tp. acidophilum* E2 catalytic domain has been solved by X-ray crystallography to 4 Å, and verified by small-angle X-ray scattering and equilibrium analytical ultracentrifugation. The *Tp. acidophilum* E2 forms a novel arrangement comprising 42 monomers, arranged into a cage-like spheroid with both square and pentagonal faces. The pentagonal faces of the 42-mer have similar dimensions to the pentagonal faces of the dodecahedral 60-mer; however, the square faces of the 42-mer are larger than those of the cubic 24-mer. The evolutionary origins of this 42-mer structure are, as yet, unknown; solving the structures of other archaeal E2s, particularly the *Picrophilus torridus* E2, may provide some insights into the diversity of archaeal E2 structures.

The novel 42-mer E2 assembly does not conform to the strict rules thought to govern the assembly of E2 structures; Izard et al. (1999) describes how a cubic 24-mer can be converted into a dodecahedral 60-mer based using Euclidean principles, and the rules of quasi-equivalence. These two structures contain identical trimeric subunits which all inhabit quasi-equivalent environments, unlike the *Tp. acidophilum* E2 trimers; the environment surrounding *Tp. acidophilum* E2 trimers and thus the ability to form different inter-trimer interactions varies depending upon if it is part of a square or pentagonal face. The differences between the *Tp. acidophilum* E2 trimer-trimer interfaces extend beyond the constraint of quasi-equivalence, as some bonds are not merely distorted by up to 5° rather they may only be present at some faces and not other. Although this structure challenges some of the symmetry principles previously thought to govern protein assembly, it does support Casper and Klug hypothesis, that biological molecules do not have to be perfectly symmetrical structures, rather subunits may be arranged into the most stable assembly possible (Caspar and Klug, 1962). Virus capsids are another example of protein assembly that do not always display perfect symmetry, in that it is not possible to arrange more than 60 subunits symmetrically; the Bacteriophage fr capsid is an example of a quasi-symmetrical structure comprising 180 identical subunits (Liljas et al., 1994). If protein assemblies, like the virus capsids, are not restrained by the principles of Euclidean symmetry, then this massively increases the number of possible subunit arrangements, including shapes like the *Tp. acidophilum* E2 which comprises multiple-shaped faces.

### *9.2.2 A functional trimer multienzyme complex*

A functional trimeric *Tp. acidophilum* OADHC has been created by the removal of the Ile anchor residue at the C-terminus of the E2 polypeptide, a key residue in the formation of trimer-trimer interactions. This trimeric complex is the first reported case of a functional E2 that is capable of binding the E1 and E3 enzymes, and catalysing the 2-oxoacid oxidative decarboxylation reaction. This trimeric complex proves that it is not essential that these OADHCs form the large 24-, 42-, or 60-mer structures. However, the fact that no naturally occurring OADHCs comprising less than 24 E2s have been described, strongly suggests that the larger assemblies are advantageous. The ability to channel substrates around the complex during limited substrate conditions (Danson et al., 1978), may increase the efficiency of carbon metabolism and thus provide selective advantage to an organism.

## 9.3 The roles of the OADHCs in archaea

The role the OADHC plays in archaeal central metabolism remains unclear. However, in all species of aerobic archaea sequenced, putative OADHC genes can be identified and in all cases are accompanied by a lipoylation system. The maintenance of these genes and their essential protein motifs, throughout archaeal evolution, strongly implies that these complexes have a necessary function, as they would not be maintained if they were redundant.

The substrate specify of only two archaeal OADHCs has been determined: the *Tp. acidophilum* OADHC is specific for branched-chain 2-oxoacids and also displays some minor activity with pyruvate, and the *S. solfataricus* OADHC appears to be specific for acetoin, based on the E1 specificity. The genomes of these two organisms do not appear to contain an FOR annotated as specific for either the branched-chain 2-oxoacids or acetoin; however, it must duly be noted that substrate specificity and promiscuity of the other FORs have not determined. The OADHCs and the FORs may provide complementary activities, increasing the range of substrates on which these organisms can survive.

The role and evolutionary origins of the OADHC still require further investigation. The size of the OADHC operon, and the additional requirement for either the LipA/LipB or LlpA lipoylation system, decreases the likelihood that all aerobic archaea independently

inherited the OADHC via lateral gene transfer from a cohabiting bacterium. This, coupled with the recent discovery of a putative OADHC operon in the anaerobic methanogen, *Methanocella* sp. HZ254 (genes numbers: mtc1889-1893), may support a hypothesis that all archaea initially possessed both the FOR and OADHC genes, but these genes have not been retained by the majority of anaerobic species. Growth studies, gene knock-outs and transcriptomics, may provide some insights into the regulation and role of these complexes. Sisignano et al (2010) found that when the *Haloferax volcanii* branched-chain OADHC was knocked-out, the organism was unable to grow on isoleucine but could still grow on leucine and valine; another study by Cozen et al (2009) found that the *Pyrobaculum aerophilum* complex genes were up-regulated when the organism was cultured with oxygen or nitrate as the terminal electron acceptors compared to arsenate, or ferric iron-citrate.

A final hypothesis can be proposed that all archaea initially possessed both the FORs and OADHC genes; however, as the organisms adapted and evolved into their niche environments; the anaerobic archaea lost the OADHCs, and the aerobic species maintained the genes. The anaerobic archaea may have favoured the 2-oxoacid ferredoxin oxidoreductase, as this enzyme utilises ferredoxin rather than $NAD^+$; whereas, the aerobic archaea have the electron transport chain, which is capable of regenerating NAD+, and therefore could use either the OADHCs or FORs. The aerobic archaea may have exploited the complementary activities of the FORs and OADHCs in order to create a specialist and separate enzyme system capable of catabolising any scavenged 2-oxoacids, whilst leaving the FORs unaffected, thus providing a competitive advantage in the hostile environments that these organism inhabit.

# Chapter 10. References


AASS, H. C. (2003) Multienzyme complexes in the Archaea. M. Phil. Thesis, University of Bath.

ÆVARSSON, A., SEGER, K., TURLEY, S., SOKATCH, J. R. & HOL, W. G. J. (1999) Crystal structure of 2-oxoisovalerate and dehydrogenase and the architecture of 2-oxo acid dehydrogenase multienzyme complexes. *Nature Structural Biology,* 6**,** 785-792.

AL-MAILEM, D. M., HOUGH, D. W. & DANSON, M. J. (2008) The 2-oxoacid dehydrogenase multienzyme complex of Haloferax volcanii. *Extremophiles,* 12**,** 89-96.

ALBERS, S. V., JONUSCHEIT, M., DINKELAKER, S., URICH, T., KLETZIN, A., TAMPE, R., DRIESSEN, A. J. M. & SCHLEPER, C. (2006) Production of recombinant and tagged proteins in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Applied and environmental microbiology,* 72**,** 102-111.

ALLEN, M. D., BROADHURST, R. W., SOLOMON, R. G. & PERHAM, R. N. (2005) Interaction of the E2 and E3 components of the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*. *FEBS Journal,* 272**,** 259-268.

ARJUNAN, P., NEMERIA, N., BRUNSKILL, A., CHANDRASEKHAR, K., SAX, M., YAN, Y., JORDAN, F., GUEST, J. R. & FUREY, W. (2002) Structure of the pyruvate dehydrogenase multienzyme complex E1 component from *Escherichia coli* at 1.85Å resolution. *Biochemistry,* 41**,** 5213-5221.

BAILEY, S. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D,* 50**,** 760-763.

BARNS, S. M., DELWICHE, C. F., PALMER, J. D. & PACE, N. R. (1996) Perspectives on Archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proceedings of the National Academy of Sciences of the United States of America,* 93**,** 9188-9193.

BELL, G. S., RUSSELL, R. J. M., CONNARIS, H., HOUGH, D. W., DANSON, M. J. & TAYLOR, G. L. (2002) Stepwise adaptations of citrate synthase to survival at life's extremes. *European Journal of Biochemistry,* 269**,** 6250-6260.

BLOMQVIST, K., NIKKOLA, M., LEHTOVAARA, P., SUIHKO, M. L., AIRAKSINEN, U., STRÄBY, K. B., KNOWLES, J. K. & PENTTILÄ, M. E. (1993) Characterization of the genes of the 2, 3-butanediol operons from Klebsiella terrigena and Enterobacter aerogenes. *Journal of bacteriology,* 175**,** 1392-1404.

BOUCHER, Y., KAMEKURA, M. & DOOLITTLE, W. F. (2004) Origins and evolution of isoprenoid lipid biosynthesis in archaea. *Molecular microbiology,* 52**,** 515-527.

BRAUTIGAM, C. A., WYNN, R. M., CHUANG, J. L. & CHUANG, D. T. (2009) Subunit and catalytic component stoichiometries of an *in vitro* reconstituted human pyruvate dehydrogenase complex. *Journal of Biological Chemistry,* 284**,** 13086-13098.

BROCK, T. D., BROCK, K. M., BELLY, R. T. & WEISS, R. L. (1972) *Sulfolobus*: A new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Archives of Microbiology,* 84**,** 54-68.

CAIVANO, A., DORIA-ROSE, N. A., BUELOW, B., SARTORIUS, R., TROVATO, M., D'APICE, L., DOMINGO, G. J., SUTTON, W. F., HAIGWOOD, N. L. & DE BERARDINIS, P. (2010) HIV-1 Gag p17 presented as virus-like particles on the E2 scaffold from *Geobacillus stearothermophilus* induces sustained humoral and cellular immune responses in the absence of IFNγ production by CD4+ T cells. *Virology,* 407**,** 296-305.

CASPAR, D. L. (1980) Movement and self-control in protein assemblies. Quasi-equivalence revisited. *Biophysical journal,* 32**,** 103-138.

CASPAR, D. L. & KLUG, A. (1962) Physical principles in the construction of regular viruses. Cold Spring Harbor Laboratory Press.

CHABRIÈRE, E., CHARON, M.-H., VOLBEDA, A., PIEULLE, L., HATCHIKIAN, E. C. & FONTECILLA-CAMPS, J.-C. (1999) Crystal structures of the key anaerobic enzyme pyruvate:ferredoxin oxidoreductase, free and in complex with pyruvate. *Nature Structural & Molecular Biology,* 6**,** 182-190.

CHAUHAN, H. J., DOMINGO, G. J., JUNG, H.-I. & PERHAM, R. N. (2000) Sites of limited proteolysis in the pyruvate decarboxylase component of the pyruvate dehydrogenase multienzyme complex of Bacillus stearothermophilus and their role in catalysis. *European Journal of Biochemistry,* 267**,** 7158-7169.

COBUCCI-PONZANO, B., CONTE, F., BENELLI, D., LONDEI, P., FLAGIELLO, A., MONTI, M., PUCCI, P., ROSSI, M. & MORACCI, M. (2006) The gene of an archaeal α-fucosidase is expressed by translational frameshifting. *Nucleic Acids Research,* 34**,** 4258-4268.

COZEN, A. E., WEIRAUCH, M. T., POLLARD, K. S., BERNICK, D. L., STUART, J. M. & LOWE, T. M. (2009) Transcriptional map of respiratory versatility in the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. *Journal of Bacteriology,* 191**,** 782-794.

CRIPPS, R. E., ELEY, K., LEAK, D. J., RUDD, B., TAYLOR, M., TODD, M., BOAKES, S., MARTIN, S. & ATKINSON, T. (2009) Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production. *Metabolic Engineering,* 11**,** 398-408.

CSŰRÖS, M. & MIKLÓS, I. (2009) Streamlining and large ancestral genomes in Archaea inferred with a phylogenetic Birth-and-Death model. *Molecular Biology and Evolution,* 26**,** 2087-2095.

DALMAU, M., LIM, S., CHEN, H. C., RUIZ, C. & WANG, S. W. (2008) Thermostability and molecular encapsulation within an engineered caged protein scaffold. *Biotechnology and Bioengineering,* 101**,** 654-664.

DALMAU, M., LIM, S. & WANG, S. W. (2009) pH-Triggered disassembly in a caged protein complex. *Biomacromolecules***,** 2306-2309.

DANIEL, R. M. & DANSON, M. J. (2010) A new understanding of how temperature affects the catalytic activity of enzymes. *Trends in Biochemical Sciences,* 35**,** 584-591.

DANIEL, R. M., DANSON, M. J., HOUGH, D. W., LEE, C. K., PETERSON, M. E. & COWAN, D. A. (2008) Enzyme stability and activity at high temperatures. IN SIDDIQUI, K. S. & THOMAS, T. (Eds.) *Protein adaptations in extremophiles.* New York, Nova Biochemical Books.

DANSON, M. J., EISENTHAL, R., HALL, S., KESSELL, S. & WILLIAMS, D. (1984) Dihydrolipoamide dehydrogenase from halophilic archaebacteria. *Biochemical Journal* 218**,** 811-818.

DANSON, M. J., FERSHT, A. R. & PERHAM, R. N. (1978) Rapid intramolecular coupling of active sites in the pyruvate dehydrogenase complex of *Escherichia coli*: mechanism for rate enhancement in a multimeric structure. *Proceedings of the National Academy of Sciences,* 75**,** 5386-5390.

DANSON, M. J., LAMBLE, H. J. & HOUGH, D. W. (2007) *Central Metabolism,* Washington, DC, ASM Press.

DARLAND, G., BROCK, T. D., SAMSONOFF, W. & CONTI, S. F. (1970) A thermophilic, acidophilic mycoplasma isolated from a coal refuse pile. *Science,* 170**,** 1416-1418.

DOMINGO, G. J., ORRU, S. & PERHAM, R. N. (2001) Multiple display of peptides and proteins on a macromolecular scaffold derived from a multienzyme complex. *Journal of Molecular Biology,* 305**,** 259-267.

DUCKWORTH, H. W., JAENICKE, R., PERHAM, R. N., WILKIE, A. O. M., FINCH, J. T. & ROBERTS, G. C. K. (1982) Limited proteolysis and proton NMR spectroscopy of *Bacillus stearothermophilus* pyruvate dehydrogenase multienzyme complex. *European Journal of Biochemistry,* 124**,** 63-69.

EISENTHAL, R., DANSON, M. J. & HOUGH, D. W. (2007) Catalytic efficiency and $k_{cat}/K_M$: a useful comparator? *Trends in Biotechnology,* 25**,** 247-249.

EMSLEY, P. & COWTAN, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography,* 60**,** 2126-2132.

FERGUSON, N., SHARPE, T. D., SCHARTAU, P. J., SATO, S., ALLEN, M. D., JOHNSON, C. M., RUTHERFORD, T. J. & FERSHT, A. R. (2005) Ultra-fast barrier-limited folding in the peripheral subunit-binding domain family. *Journal of Molecular Biology,* 353**,** 427-446.

FRANK, R. A. W., PRATAP, J. V., PEI, X. Y., PERHAM, R. N. & LUISI, B. F. (2005) The molecular origins of specificity in the assembly of a multienzyme complex. *Structure* 13**,** 1119-1130.

FRANK, R. A. W., TITMAN, C. M., PRATAP, J. V., LUISI, B. F. & PERHAM, R. N. (2004) A molecular switch and proton wire synchronize the active sites in thiamine enzymes. *Science,* 306**,** 872-876.

FRANKE, D. & SVERGUN, D. I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of Applied Crystallography,* 42**,** 342-346.

FRIES, M., JUNG, H.-I. & PERHAM, R. N. (2003) Reaction mechanism of the heterotetrameric ($\alpha_2\beta_2$) E1 component of 2-oxo acid dehydrogenase multienzyme complexes *Biochemistry,* 42**,** 6996-7002.

FRIES, M., STOTT, K. M., REYNOLDS, S. & PERHAM, R. N. (2007) Distinct modes of recognition of the lipoyl domain as substrate by the E1 and E3 components of the pyruvate dehydrogenase multienzyme complex. *Journal of Molecular Biology,* 366**,** 132-139.

FUTTERER, O., ANGELOV, A., LIESEGANG, H., GOTTSCHALK, G., SCHLEPER, C., SCHEPERS, B., DOCK, C., ANTRANIKIAN, G. & LIEBL, W. (2004) Genome sequence of *Picrophilus torridus* and its implications for life around pH 0. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 9091.

GASTEIGER, E., GATTIKER, A., HOOGLAND, C., IVANYI, I., APPEL, R. D. & BAIROCH, A. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research,* 31**,** 3784-3788.

GUEST, J. R. (1987) Functional implications of structural homologies between chloramphenicol acetyltransferase and dihydrolipoamide acetyltransferase. *FEMS microbiology letters,* 44**,** 417-422.

HAWKINS, C. F., BORGES, A. & PERHAM, R. N. (1989) A common structural motif in thiamin pyrophosphate-binding enzymes. *FEBS Letters,* 255**,** 77-82.

HEATH, C. (2006) A branched-chain 2-oxoacid dehydrogenase multienzyme complex from *Thermoplasma acidophilum*. Ph.D. Thesis, University of Bath.

HEATH, C., JEFFRIES, A. C., HOUGH, D. W. & DANSON, M. J. (2004) Discovery of the catalytic function of a putative 2-oxoacid dehydrogenase multienzyme complex in the thermophilic archaeon *Thermoplasma acidophilum*. *FEBS Letters,* 577**,** 523-527.

HEATH, C., POSNER, M. G., AASS, H. C., UPADHYAY, A., SCOTT, D. J., HOUGH, D. W. & DANSON, M. J. (2007) The 2-oxoacid dehydrogenase multi-enzyme complex of the archaeon *Thermoplasma acidophilum* - recombinant expression, assembly and characterization. *FEBS Journal,* 274**,** 5406-5415.

HIPPS, D. S., PACKMAN, L. C., ALLEN, M. D., FULLER, C., SAKAGUCHI, K., APPELLA, E. & PERHAM, R. N. (1994) The peripheral subunit-binding domain of the dihydrolipoyl acetyltransferase component of the pyruvate dehydrogenase complex of Bacillus stearothermophilus: preparation and characterization of its binding to the dihydrolipoyl dehydrogenase component. *Biochemical Journal,* 297**,** 137.

HIROMASA, Y., FUJISAWA, T., ASO, Y. & ROCHE, T. E. (2004) Organization of the cores of the mammalian pyruvate dehydrogenase complex formed by E2 and E2 plus the E3-binding protein and their capacities to bind the E1 and E3 Components. *Journal of Biological Chemistry,* 279**,** 6921-6933.

HOLM, L. & PARK, J. (2000) DaliLite workbench for protein structure comparison. *Bioinformatics,* 16**,** 566-567.

HOUGH, D. W. & DANSON, M. J. (1999) Extremozymes. *Current opinion in chemical biology,* 3**,** 39-46.

HUBER, H., HOHN, M. J., RACHEL, R., FUCHS, T., WIMMER, V. C. & STETTER, K. O. (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature,* 417**,** 63-67.

IZARD, T., ÆVARSSON, A., ALLEN, M. D., WESTPHAL, A. H., PERHAM, R. N., DE KOK, A. & HOL, W. G. J. (1999) Principles of quasi-equivalence and Euclidean geometry govern the assembly of cubic and dodecahedral cores of pyruvate dehydrogenase complexes. *Proceedings of the National Academy of Sciences,* 96**,** 1240.

JACQUES, D. A. & TREWHELLA, J. (2010) Small-angle scattering for structural biology - Expanding the frontier while avoiding the pitfalls. *Protein Science,* 19**,** 642-657.

JOLLEY, K. A., MADDOCKS, D. G., GYLES, S. L., MULLAN, Z., TANG, S. L., DYALL-SMITH, M. L., HOUGH, D. W. & DANSON, M. J. (2000) 2-Oxoacid dehydrogenase multienzyme complexes in the halophilic Archaea? Gene sequences and protein structural predictions. *Microbiology* 146**,** 1061.

JUNG, H.-I., COOPER, A. & PERHAM, R., N. (2003) Interactions of the peripheral subunit-binding domain of the dihydrolipoyl acetyltransferase component in the assembly of the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*. *European Journal of Biochemistry,* 270**,** 4488-4496.

JUNG, H.-I., COOPER, A. & PERHAM, R. N. (2002) Identification of key amino acid residues in the assembly of enzymes into the pyruvate dehydrogenase complex of *Bacillus stearothermophilus*: A kinetic and thermodynamic analysis. *Biochemistry,* 41**,** 10446-10453.

JURGENS, G. (2002) Molecular phylogeny of Archaea in boreal forest soil, freshwater and temperate estuarine sediment. *Department of Applied Chemistry and Microbiology* University of Helsinki.

KALIA, Y. N., BROCKLEHURST, S. M., HIPPS, D. S., APPELLA, E., SAKAGUCHI, K. & PERHAM, R. N. (1993) The high-resolution structure of the peripheral subunit-binding domain of dihydrolipoamide acetyltransferase from the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*. *Journal of Molecular Biology,* 230**,** 323-341.

KARNER, M. B., DELONG, E. F. & KARL, D. M. (2001) Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature,* 409**,** 507-510.

KATO, M., WYNN, R. M., CHUANG, J. L., BRAUTIGAM, C. A., CUSTORIO, M. & CHUANG, D. T. (2006) A synchronized substrate-gating mechanism revealed by cubic-core structure of the bovine branched-chain $\alpha$-ketoacid dehydrogenase complex. *EMBO J,* 25**,** 5983-5994.

KERSCHER, L., NOWITZKI, S. & OESTERHELT, D. (1982) Thermoacidophilic archaebacteria contain bacterial-type ferredoxins acting as electron acceptors of 2-oxoacid: ferredoxin oxidoreductases. *European Journal of Biochemistry,* 128**,** 223-230.

KERSCHER, L. & OESTERHELT, D. (1981) The catalytic mechanism of 2-oxoacid: Ferredoxin oxidoreductases from *Halobacterium halobium*. *European Journal of Biochemistry,* 116**,** 595-600.

KERSCHER, L. & OESTERHELT, D. (1982) Pyruvate : ferredoxin oxidoreductase - new findings on an ancient enzyme. *Trends in Biochemical Sciences,* 7**,** 371-374.

KNAPP, J. E., CARROLL, D., LAWSON, J. E., ERNST, S. R., REED, L. J. & HACKERT, M. L. (2000) Expression, purification, and structural analysis of the trimeric form of the catalytic domain of the *Escherichia coli* dihydrolipoamide succinyltransferase. *Protein Science,* 9**,** 37-48.

KNAPP, J. E., MITCHELL, D. T., YAZDI, M. A., ERNST, S. R., REED, L. J. & HACKERT, M. L. (1998) Crystal structure of the truncated cubic core component of the *Escherichia coli* 2-oxoglutarate dehydrogenase multienzyme complex. *Journal of Molecular Biology,* 280**,** 655-668.

KOBAYASHI, K., KUSAKA, K., TAKAHASHI, T. & SATO, K. (2005) Method for the simultaneous assay of diacetyl and acetoin in the presence of $\alpha$–acetolactate: Application in determining the kinetic parameters for the decomposition of $\alpha$–acetolactate. *Journal of Bioscience and Bioengineering,* 99**,** 502-507.

KOIKE, K., SUEMATSU, T. & EHARA, M. (2000) Cloning, overexpression and mutagenesis of cDNA encoding dihydrolipoamide succinyltransferase component of the porcine 2-oxoglutarate dehydrogenase complex. *European Journal of Biochemistry,* 267**,** 3005-3016.

KONAREV, P. V., VOLKOV, V. V., SOKOLOVA, A. V., KOCH, M. H. J. & SVERGUN, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography,* 36**,** 1277-1282.

LENGYEL, J. S., STOTT, K. M., WU, X., BROOKS, B. R., BALBO, A., SCHUCK, P., PERHAM, R. N., SUBRAMANIAM, S. & MILNE, J. L. S. (2008) Extended polypeptide linkers establish the spatial architecture of a pyruvate dehydrogenase multienzyme complex. *Structure,* 16**,** 93-103.

LESSARD, I. A. & PERHAM, R. N. (1995) Interaction of component enzymes with the peripheral subunit-binding domain of the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*: stoichiometry and specificity in self-assembly. *Biochemical Journal,* 306**,** 727.

LI, W. F., ZHOU, X. X. & LU, P. (2005) Structural features of thermozymes. *Biotechnology Advances,* 23**,** 271-281.

LILJAS, L., FRIDBORG, K., VALEGARD, K., BUNDULE, M. & PUMPENS, P. (1994) Crystal Structure of Bacteriophage fr Capsids at 3.5Å Resolution. *Journal of Molecular Biology,* 244**,** 279-290.

LONG, F., VAGIN, A. A., YOUNG, P. & MURSHUDOV, G. N. (2007) BALBES: a molecular-replacement pipeline. *Acta Crystallographica Section D: Biological Crystallography,* 64**,** 125-132.

MANDE, S. S., SARFATY, S., ALLEN, M. D., PERHAM, R. N. & HOL, W. G. J. (1996) Protein-protein interactions in the pyruvate dehydrogenase multienzyme complex: dihydrolipoamide dehydrogenase complexed with the binding domain of dihydrolipoamide acetyltransferase. *Structure,* 4**,** 277-286.

MARROTT, N. L., MARSHALL, J. J. T., SVERGUN, D. I., CRENNELL, S. J., HOUGH, D. W., DANSON, M. J. & VAN DEN ELSEN, J. M. H. (2012) The catalytic core of an archaeal 2-oxoacid dehydrogenase multienzyme complex is a 42-mer protein assembly. *FEBS Journal,* 279**,** 713-723.

MATTEVI, A., OBMOLOVA, G., KALK, K. H., TEPLYAKOV, A. & HOL, W. G. J. (1993) Crystallographic analysis of substrate binding and catalysis in dihydrolipoyl transacetylase (E2p). *Biochemistry,* 32**,** 3887-3901.

MATTEVI, A., OBMOLOVA, G., SCHULZE, E., KALK, K. H., WESTPHAL, A. H., KOK, A. D. & HOL, W. G. J. (1992) Atomic structure of the cubic core of the pyruvate dehydrogenase multienzyme complex. *Science,* 255**,** 1544-1550.

MATTEVI, A., SCHIERBEEK, A. J. & HOL, W. G. (1991a) Refined crystal structure of lipoamide dehydrogenase from *Azotobacter vinelandii* at 2.2Å resolution. A comparison with the structure of glutathione reductase. *Journal of Molecular Biology,* 220**,** 975.

MATTEVI, A., SCHIERBEEK, A. J. & HOL, W. G. J. (1991b) Refined crystal structure of lipoamide dehydrogenase from Azotobacter vinelandii at 2.2 Å resolution : A comparison with the structure of glutathione reductase. *Journal of Molecular Biology,* 220**,** 975-994.

MCMANUS, E., LUISI, B. F. & PERHAM, R. N. (2006) Structure of a Putative Lipoate Protein Ligase from *Thermoplasma acidophilum* and the Mechanism of Target Selection for Post-translational Modification. *Journal of Molecular Biology,* 356**,** 625-637.

MORRIS, T. W., REED, K. E. & CRONAN, J. E. (1994) Identification of the gene encoding lipoate-protein ligase A of *Escherichia coli*. Molecular cloning and characterization of the lplA gene and gene product. *Journal of Biological Chemistry,* 269**,** 16091-100.

NEMERIA, N. S., ARJUNAN, P., CHANDRASEKHAR, K., MOSSAD, M., TITTMANN, K., FUREY, W. & JORDAN, F. (2010) Communication between thiamin cofactors in the *Escherichia coli* pyruvate dehydrogenase complex E1 component active centers. *Journal of Biological Chemistry,* 285**,** 11197-11209.

OGATA, H., GOTO, S., SATO, K., FUJIBUCHI, W., BONO, H. & KANEHISA, M. (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research,* 27**,** 29-34.

OZAWA, Y., NAKAMURA, T., KAMATA, N., YASUJIMA, D., URUSHIYAMA, A., YAMAKURA, F., OHMORI, D. & IMAI, T. (2005) *Thermococcus profundus* 2-ketoisovalerate ferredoxin oxidoreductase, a key enzyme in the archaeal energy-producing amino acid metabolic pathway. *Journal of Biochemistry,* 137**,** 101-107.

PACKMAN, L. C., PERHAM, R. N. & ROBERTS, G. C. (1984) Domain structure and 1H-nmr spectroscopy of the pyruvate dehydrogenase complex of *Bacillus stearothermophilus*. *Biochemical Journal,* 217**,** 219.

PARK, Y., YOO, C., CHOI, S. Y. & LEE, H. (2006) Purifications and characterizations of a ferredoxin and its related 2-oxoacid: Ferredoxin oxidoreductase from the hyperthermophilic Archaeon, *Sulfolobus solfataricus* P1. *Journal of Biochemistry and Molecular Biology,* 39**,** 46.

PAYNE, K. A. P., HOUGH, D. W. & DANSON, M. J. (2010) Discovery of a putative acetoin dehydrogenase complex in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *FEBS letters,* 584**,** 1231-1234.

PEI, X. Y., TITMAN, C. M., FRANK, R. A. W., LEEPER, F. J. & LUISI, B. F. (2008) Snapshots of catalysis in the E1 subunit of the pyruvate dehydrogenase multienzyme complex. *Structure,* 16**,** 1860-1872.

PENG, T., LEE, H. & LIM, S. (2012) Isolating a trimer intermediate in the self-assembly of E2 protein cage. *Biomacromolecules,* 13**,** 699-705.

PERHAM, R. N. (1991) Domains, motifs, and linkers in 2-oxo acid dehydrogenase multienzyme complexes: a paradigm in the design of a multifunctional protein. *Biochemistry,* 30**,** 8501-8512.

PERHAM, R. N. (2000) Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. *Annual review of biochemistry,* 69**,** 961-1004.

PETOUKHOV, M. V. & SVERGUN, D. I. (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophysical journal,* 89**,** 1237-1250.

POSNER, M. G. (2009) Lipoylation and assembly of a 2-oxoacid dehydrogenase multienzyme complex from *Thermoplasma acidophilum*. *Biology and Biochemistry.* Ph.D. Thesis, University of Bath.

POSNER, M. G., UPADHYAY, A., BAGBY, S., HOUGH, D. W. & DANSON, M. J. (2009) A unique lipoylation system in the Archaea. Lipoylation in *Thermoplasma acidophilum* requires two proteins. *The FEBS journal,* 276**,** 4012.

PRATT, K. J., CARLES, C., CARNE, T. J., DANSON, M. J. & STEVENSON, K. J. (1989) Detection of bacterial lipoic acid. A modified gas-chromatographic-mass-spectrometric procedure. *Biochemical Journal,* 258**,** 749.

PRIEFERT, H., HEIN, S., KRÜGER, N., ZEH, K., SCHMIDT, B. & STEINBÜCHEL, A. (1991) Identification and molecular characterization of the *Alcaligenes eutrophus* H16 aco operon genes involved in acetoin catabolism. *Journal of Bacteriology,* 173**,** 4056-4071.

RADESTOCK, S. & GOHLKE, H. (2010) Protein rigidity and thermophilic adaptation. *Proteins: Structure, Function, and Bioinformatics,* 79**,** 1089-1108.

RECHE, P. & PERHAM, R. N. (1999) Structure and selectivity in post-translational modification: attaching the biotinylate-lysine and lipoylated-lysine swinging arms in multifunctional enzymes. *The EMBO journal,* 18**,** 2673-2682.

REED, L. J., KOIKE, M., LEVITCH, M. E. & LEACH, F. R. (1958) Studies on the nature and reactions of protein-bound lipoic acid. *Journal of Biological Chemistry,* 232**,** 143-158.

RHODES, G. (1993) *Crystallography made crystal clear: a guide for users of macromolecular models*, Academic Press Inc. California.

ROESSLE, M. W., KLAERING, R., RISTAU, U., ROBRAHN, B., JAHN, D., GEHRMANN, T., KONAREV, P., ROUND, A., FIEDLER, S. & HERMES, C. (2007) Upgrade of the small-angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg. *Journal of Applied Crystallography,* 40**,** s190-s194.

ROUND, A. R., FRANKE, D., MORITZ, S., HUCHLER, R., FRITSCHE, M., MALTHAN, D., KLAERING, R., SVERGUN, D. I. & ROESSLE, M. (2008) Automated sample-changing robot for solution scattering experiments at the EMBL Hamburg SAXS station X33. *Journal of Applied Crystallography,* 41**,** 913-917.

RUEPP, A., GRAML, W., SANTOS-MARTINEZ, M.-L., KORETKE, K. K., VOLKER, C., MEWES, H. W., FRISHMAN, D., STOCKER, S., LUPAS, A. N. & BAUMEISTER, W. (2000) The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature,* 407**,** 508-513.

SEARCY, D. G. (1976) *Thermoplasma acidophilum*: Intracellular pH and potassium concentration. *Biochimica et Biophysica Acta (BBA) - General Subjects,* 451**,** 278-286.

SHARPE, T. D., FERGUSON, N., JOHNSON, C. M. & FERSHT, A. R. (2008) Conservation of transition state structure in fast folding peripheral subunit-binding domains. *Journal of Molecular Biology,* 383**,** 224-237.

SISIGNANO, M., MORBITZER, D., GÄGENS, J., OLDIGES, M. & SOPPA, J. (2010) A 2-oxoacid dehydrogenase complex of *Haloferax volcanii* is essential for growth on isoleucine but not on other branched-chain amino acids. *Microbiology,* 156**,** 521-529.

SMITH, L. D., BUNGARD, S. J., DANSON, M. J. & HOUGH, D. W. (1987) Dihydrolipoamide dehydrogenase from the thermoacidophilic Archaebacterium *Thermoplasma acidophilum*. *Biochemical Society Transactions,* 15**,** 1097 - 1097.

SPECTOR, S., KUHLMAN, B., FAIRMAN, R., WONG, E., BOICE, J. A. & RALEIGH, D. P. (1998) Cooperative folding of a protein mini domain: the peripheral subunit-binding domain of the pyruvate dehydrogenase multienzyme complex. *Journal of Molecular Biology,* 276**,** 479-489.

SRERE, P. A. & KOSICKI, G. W. (1961) The purification of citrate-condensing enzyme. *The Journal of biological chemistry,* 236**,** 2557.

SUN, N., BECK, F., KNISPEL, R. W., SIEDLER, F., SCHEFFER, B., NICKELL, S., BAUMEISTER, W. & NAGY, I. (2007) Proteomics analysis of *Thermoplasma acidophilum* with a focus on protein complexes. *Molecular & Cellular Proteomics,* 6**,** 492.

SUN, N., PAN, C., NICKELL, S., MANN, M., BAUMEISTER, W. & NAGY, I. N. (2011) Quantitative proteome and transcriptome analysis of the Archaeon

*Thermoplasma acidophilum* cultured under aerobic and anaerobic conditions. *Journal of Proteome Research,* 9**,** 4839-4850.

SVERGUN, D., BARBERATO, C. & KOCH, M. H. J. (1995) CRYSOL-a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography,* 28**,** 768-773.

SVERGUN, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography,* 25**,** 495-503.

SVERGUN, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical journal,* 76**,** 2879.

THOMPSON, M. J. & EISENBERG, D. (1999) Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. *Journal of Molecular Biology,* 290**,** 595-604.

THORPE, C. & WILLIAMS JR, C. H. (1981) Lipoamide dehydrogenase from pig heart. Pyridine nucleotide induced changes in monoalkylated two-electron reduced enzyme. *Biochemistry,* 20**,** 1507-1513.

VIEILLE, C. & ZEIKUS, G. J. (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiology and Molecular Biology Reviews,* 65**,** 1-43.

VOLKOV, V. V. & SVERGUN, D. I. (2003) Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography,* 36**,** 860-864.

WINKER, S. & WOESE, C. R. (1991) A Definition of the Domains, Archaea, Bacteria, and Eucarya in Terms of Small Subunit Ribosomal RNA Characteristics. *Systematic and applied microbiology,* 14**,** 305-310.

YU, X., HIROMASA, Y., TSEN, H., STOOPS, J. K., ROCHE, T. E. & ZHOU, Z. H. (2008) Structures of the human pyruvate dehydrogenase complex cores: a highly conserved catalytic center with flexible N-terminal domains. *Structure,* 16**,** 104-114.

# Chapter 11: Appendixes

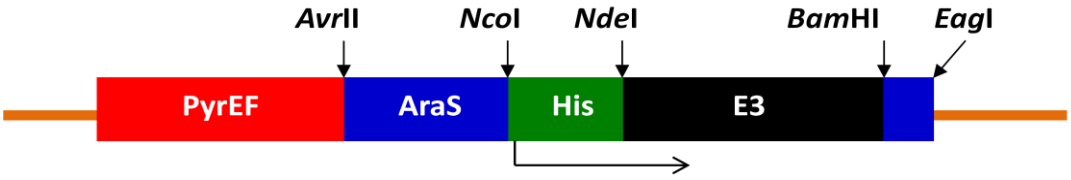
## Appendix 1 E2 and E2 catalytic domain sequence

The *Tp. acidophilum* E2 sequence is shown below with the sequence complementary to the E2 catalytic domain shown in italics and highlighted in grey.

>*Tp. acidophilum* E2
ATGTACGAATTCAAACTGCCAGACATAGGTGAGGGGGTAACAGAGGGCGAGATAGTCAGAT
GGGATGTCAAGGAAGGCGACATGGTGGAGAAGGATCAGGACCTTGTTGAGGTTATGACGGA
CAAGGTCACAGTGAAGATACCGTCGCCCGTCAGGGGAAAGATAGTGAAGATACTCTACAGG
GAGGGCCAGGTTGTCCCTGTGGGTTCTACCCTCCTGCAGATCGATACCGGTGAAGAGGCTC
CGGTACAGCAGCCTGCCGGCCGTGCCGAAAGCACGGTTCAGGTTGCAGAGGTCAAGCAGGT
GCCCCTGCCGGAGGTATCCGGGCATGTGCTTGCAAGCCCTGCAGTTAGGAGAATAGCGAGG
GAGAACGGAATAGACCTCTCGAAGGTGGGAGGCACCGGAGAGGGTGGGCGCGTGACGCTTG
ACGATCTGGAGAGGTACATGAAGTCGCCTGCACCATCGCCAGCCCCATCCGCTGGAAAAGC
GGAGGCCGTGCACACGGCGCCGCAGATCCCGGCCCAGAAGCCT*GCACCTGGGAGAGAAGAG*
*ATCCTGG*AGATGCATGGCCTGAGGCGGATAATATTCGACAAGATGACGAAGGCAAAGCAGA
TCATGCCCCATTTCACGGTGATGGAGGAAGTCGACGTCACATCCATGGTTTCGATTCTGGA
TTCCGCGAAGGCCAGGAACAGGAAGGTCACCGTCACGGGCTTTCTGGCAAGGATCGTGCCA
TCTATCCTGAAGCAGTATCCATACCTCAACGCAATTTACGATGAAACGCGCAGAGTTTACA
TACTGAAGAAATACTACAACATAGGCATAGCGGTTGACACGCCGGACGGCCTGAACGTCTT
CGTCATAAAGGATGCTGATAGAAAGAGCATGGTGGAGATATCTGCAGAGATATCGGACAAG
GCAAGCAGGGCACGCGAGAACAAGCTGCAGCTGGATGAAGTCCAGGATTCAACGTTCACCA
TAACCAACGTTGGCACAATAGGGGGCATAATGTCCACGCCAATAATCAACTATCCTGAGGT
CGCAATACTGGGTGTGCACCGGATCCTGGAAAGAGAGGGGCGGAAATACATGTACCTATCG
CTGTCCTGTGACCACAGGCTCATAGACGGAGCGGTTGCAACCAGGTTCATCGTGGATCTGA
AAAAGGTTATTGAGGAT*CCGAACGCTATAATCTACGAGATCTGA*

## Appendix 2: *Sulfolobus solfataricus* E3-pMJ0503 sequence

The pMJ0503 viral vector (orange) containing the *S. solfataricus* E3 gene is shown in black, with the start ATG underlined, the his-tag (His) originally from pET28a is shown in green, the arabinose promoter (AraS) is shown in blue, and the marker gene PyrEF that complements the uracil auxotrophic mutant *S. solfataricus* strain is shown in red. The following restriction sites have been highlighted in the sequence: AvrII, NcoI, *Bam*HI and *Eag*I.




**>pMJ0503-E3**

CTGATACCATTAGGATGCTAAACAACTATTCAAACTGTATTGTAATGAGACATAAGTTTGA
TGGGGCAGCATTATTCCCTAGGGCACCATATGTTTAGAGATGAAGCTTAGAAGATCTTAGA
TAATCTGAGTTTGATCTTTTATGTGCATTGTGGTCATGTTGAATTTTCACGATCATTTAAG
GACTCCCATAAACATAAATTATGTATCAAAACATTAATTGAAATATAGATAATAGTTATAT
TATAGTTATTTTTAGAAAAACATCCAATATGTTAACAAAACGTCTTTTACGGAAATATATA
AATGTTAAACAAGTTAGGTATACTATTTATAAAATAGTTAGGTCATAAAAGTACCCGAGAC
C**ATGG**GCAGCAGCCATCATCATCATCATCAC**AGCAGCGGCCTGGTGCCGCGCGGCAGCCAT**
**ATGAAAGTTGTAATAATTGGTGCAGGTCCAGCAGGAGTTTACTCAGCATTAACCTTATCCA**
**AACACGCAAAAGTTACACTTATAGAAAGGGAAGAAAAACTAGGGGGTACTTGCGTACTTTA**
**CGGATGTATACCCACAAAATCCATACTAAGCCAGCTTATCATATCTAGGCAAGCATCCAAT**
**ATGAGTTTAAATACGCTCAGAGAATACGCATTGACTTCAATAAATACTATAAGTAAGAGTT**
**TAGAACACTTGCTAAATAGTCATGGGATAGAGGTTATTCATGCTAATGGCTTTCTAAGATC**
**GTCTATGGTTCACGCTTCAAATACTTCACTTGCGGCTGATAAGGTTTTAGTTTCAACCGGA**
**ACCAGAAGAGAAAGATTAGGAAAGGTAAAATTCACAGAGGATTTAGCCTATACTAATGAGG**
**ATTATAATAAGGTGGTTATAGTAGGAGGAGATGCAGGTGGGATAGAATTAGGTTGGATGAT**
**GAAGAAACTGGGTAAGGAGGTACATCTTATAGACAAGAACGATTTACTTTTACAAAACATT**
**GATAGGACCCTCTCTGAGATAGTTACAAATTTTCTAAGTCAGATAGGGATTAAATTATACT**
**TAGGTAAAAAGGTCTCTAAAATTGACGAAACGTCTGTTACTCTTGAAGATAATCAAAAGAT**
**CTCTGGCGATGCTGTATTCGTAACCTTCGGTCGTAAACCAAATATTGAAGGATTCGAGGAA**
**ATTCCCCACGAGAAGTACATATACGTTGACGAATATTTGAGAACTCAAATTCCTAACATCT**
**ATGCAGCAGGGGATATAATAGGTACGTTTACTGCACATGAAGCCATATATGCTGGTATTAT**
**AGCTGCGAAGAACATGTTAGGCGAAAAAAGGGAGTTCATCGTAGAAGGAATACCGAAGGTT**
**ATATACATTTACCCACAAATAGCCTATGTAGGTACTACTAATGGTAATTGTGTTACTTTTA**
**ACACGTTAAATTTAACTAGAACAATAGTAGAAAGGGAAAGTGAAGGATTCTTAAAGATATG**
**CGAGAGAGATAACAGAGTAATTGGGGCGGTAGCGTTCATGCCCTATGCAGAAGACGTAATT**
**TCTCTTATTTCGGTATTAATACGATACCAAATAAGTTTGAAGGATACGATAGATCTTGTTA**
**TGCCTCATCCTTCTTATCTGGAAGCAATTACTGAGGCTTTAAATATGTTACAAAGTAAAGG**
**GTAA**GGATCCGGATGGAGTCATCCACAATTTGAGAAGCATCACCATCATCACCATCATCAC
CATCATTGAGGGCCCCACTTTCTCAAGTCTCACTATACCAAATGAGTTTTCTTTTAATCTT
ATTCTAATCTCATTTTCATTAGATTGCCGGCCGCCACCGCGGTGGAGCTCGAATTCGTAAT
CATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATAC

## Appendix 3: *Picrophilus torridus* OADHC genes

The bases at the end and beginning of genes which overlap with the adjacent gene have been underlined.

**> *P. torridus* LipA**

ATGAATGCAGGTCCAATAAACTACAAGGTTAAACTGCCCTCGGGGGAAAGATATACTTTTA
TCAAGAGTACATTATCGGCAAGAAATCTTTACACTGTCTGCGAGGAGGCGCACTGCCCAAA
CATAGCCGAATGCTGGGAAAGCGGTACTGCAACCTTTATGATAATGGGTTCAAACTGCTCC
AGGGGATGCAGGTTCTGTGCCGTCACCCATGGAAGGATGCTGCCCCTGGATCCAATGGAGC
CTGAGAAGGTTTACGAGTCAGTAAAAATGATGAATCTTGATTATGTTGTTATTACATCAGT
AGATCGTGATGATCTTCCGGATAAGGGTTCATCGCACTTCGCAGCGGTAATAAGAAGGTTA
AAAGACCTTAAAATAAAAATAGAGGTTTTGATCCCGGACTTCAGCGGCGTTCATAAATTTA
TTGACAAAATTATAGATGAAAGGCCTGATGTGATAGCACATAACATTGAAACGGTTCGCAG
GCTTACAAAAACTGTCAGGGATCCAAGGGCCGGATATGATCAGAGTCTTAATGTTCTAAGG
TATGTAAAGTCCAGGTCAAACATCATAACAAAATCGTCGATAATGCTTGGCCTTGGAGAGA
CCGATGATGAGGTTATCGAAACACTGCACGACCTGCATGATGCTGGTGTTGACATTGTAAC
GATTGGACAATACCTCAGACCAACAAAGAAGCAGCTTGAGGTTAAGGAGTATTCCCCAATG
GAAAGATTTAAAAATCTTGAGGAGTCCGCCTATAGCATAGGTTTTTCCTTTGTGGCCTCGG
GTCCGCTTGTAAGGACATCTTACAGGGCGGCTGAGGCTTTTGTCAAGGGGGGTTTTAAAAA
TGATTGA

**>*P. torridus* E1α**

ATGATTGAGGAAGATATTAGTAAAGAGGATATTATTAGTGCATACAGAAACATGGTTTTGG
AAAGGTTTCTTGATAAAAAACTGCTTGGAATAAACAGACAGGGATTTTTGCCATTTTATAT
ACCAAACATTGGACATGAGGCTTTGCATGCCGCGATAGGCATGGCCATAAGGGATGATGAT
TTCTTCTATCCATATTACAGGGACCTTGGATCGGACATAGCCAGGGTTGGACTTGATTTTG
TGCTTGCCCAGATGTTCTCAACGGAAATGGACAACGAGCTTGGGCGTGATATGCCGTTGCA
TATATCAAATAAGGCCAAAAGGGTTGGCCCGGTTATAACCACCGTCGGCGGGCATCTTATG
GCTGCAACCGGTGTTGCATACTCATACAAATATCAGAAGAAGCCGGGCATAGTTATTACAA
CGTTTGGTGATGGTGCAACATCAACGCCTGATTTCCATGTTTCAATGAATTTCGCAGCGGT
TTATTCGCTCCCGCTACTTTTTATATGTGAGAACAACCAGTGGGCGATATCATATCCTGTT
GAGGAGCAAACAAAGGTTGAGATATCAAAAAAGGCAGAGGCCTATGGTTTTACAGGCATAA
AGATCGATGGCAATAACTTCATAGAGGCATATCACGCAATAAGAAATGCCATTAAAGATGT
TGAAAAAAATAAAATGCCATTGTTAATAGATGCTGTAACATACAGAATGGGGCCGCATACA
ACGGCGGATGATCCAAATAAATACAGAAAAACCATAATAAATGAGGGTGATCCTCTTGATC
CACTGTCAATCATTGAGGATGATATAAAAAAGATGAAGATACTCAACGATGAGGAGATATC
AAATATAAAAAACGAGATAAATAACATGGTATCAAAGGAGGTTGAAAGATACGAGAAGATG
AACAAGCCTGGAAAGGAAACGCTGTTTAAAAACATCTATGAAAACGAGCCATGGTATATAA
CAGAGGAAAGGGGTGAGATAGAATGA

**>*P. torridus* E1β**

ATGACCGAGATGAACATGGTAAAGGCGCTTAACAGCGCACTTGATACAATGCTTGAAAGGG
ATAAAAACGTCATACTTCTTGGAGAGGATATAGCCAAGGACGGTGGTGTTTTCAGGGTTAC
CGACGGTCTTTATGCAAAGTACGGCGGGGAGCGTGTTATAAGCACACCACTCTCAGAACTT
GGCATAGTTGGCATGGGCATAGGCATGGCAATGGATGGCCTGAGGCCCGTGCCTGAGATCC
AGTTTCTGGATTTTATATACACTGCAATGGATCAAATAGTGAGCCAGATGGCAAAGATAAG
GTACAGAACAAATGGCGATTACACCCTGCCAATGGTTCTTAGAACACCATATGGCGGTGGC
GTTTCAGGCGGGCCTTATCATTCACAGAGCAGCGAGGCATACTTTGCACATACTGCAGGTC
TTGTTGTTGTCACGCCATCAAATCCATACGATGCAAAGGGCCTTTTAATATCTGCAATAGA
AAGCAATGATCCTGTTATGTTTTTGGAGCCCAAGAGGATTTACTATTCAATAAAGAACGAT
GTGCCTGACAATTATTATAAGGTTGATATAGGAAGGGCCAAAAGGATCCTTGAGGGTGATG
ATGTTACCTTAATAACATATGGCCCGATGGTTCCACTGGTAAAATCTGTTGTTCAGAAAAA

CAATGTAAATGCCGATGTCATAGATCTAATAACATTAAATCCATTTGATGTGAATTCTATA
ATAAACTCTGTTAAGAGAACGGGCCGCGCGGTAATTGTCCATGAGGCGCCAAAGATGTTTG
GTGCCGGTGCTGAAATAGCAGCAACAATAGCTGAAAAGGCCATAGATTACCTTCAGGCACC
TATTTTAAGGGTTACAGGCATGGATATACCTGTTCCTTTTATACTTGAGGATTACTACGTT
CCAAATGAGAAGAGAATTATGAATGCTATAAATAAGGTTATAAATTATTAG

**>*P. torridus* E2**

ATGTACACGTTGAAGGTTCCACCAATAGGAGAGGGCGTCAGTGAGGGCGAGATAGTAAAAT
GGAATGTAAAGGAGGGCGACACAATAGAAAAGGATCAGGAGATCGTTGAGATAATGACTGA
TAAGATAACAATAAAGATTCCATCCCCGGTCAGCGGAAAGGTATTAAAATTGATTGAGCCG
GAGGGAAAAACCGTCAAGGTCGGTGACAGCATAGCAACAATAGATTCGCAGGAGGGCAATG
AAGAAATTAACAATGAGAACAATGCCCAGGAATCAAAGGAAATTAAAATAGAAAACAAAAA
TGAGGGCAGCAATGTAAAAAACGTTGAGCTGGTAAAGGCAACACCCGCTGTGAGGGCATAC
GCAAGGCAGAAGGGAATAGATCTTTCAAATGTAAGGCCGTCCAGACCTGATGGCAGGATAA
GAAAAGAGGATATAGACAGCTATATATCAATGAAGAATAAAACTGTTCAGGAAAATGTCGA
GATACAAAACGATGAGGTATATAAACCATCAGGCATAAGAAAGATCATCTTCGACAAGATG
ACGAAATCAAAGCAGATCATACCACATTTTACAATAACAGATTTTATATCAACGGAGAATA
TTGAAAAGGCCATTGATTATTATTCAAAGAAGGGCTATGTTAGCTTTACATCATTCTTTGC
AAAGGCATGCACAATAGCATTTAAGGAGTTCCCAAAGATGAATGCCTTATATAATGATGAT
GGCACATACACAATAAAGAAGAGGTATAATATTGGAATAGCCGTTGATTCTCCATATGGAT
TAACCGTTGTTGTTGTAAAGGACGTTGATAAGAAGAGCATCTTTGAGATATCAATGGAAAT
AAGGGAGCTCGCCGAAAAGGCAAGATCAAACAAACTCGAAATGGATGATGTTAGGGACTCG
ACGTTTTCAGTAACAAACATAGGGGCAATAGGCGGAATATATTCAACACCGATAATAAACT
ATCCTGAGGTCGCAATTCTTGCTGTTAACACAAGGACAAATGCATTTATAGATGGTTCAAT
GAGAAGCGGCGTCTATGTTACACTCGCCTGCGACCACAGATTAATAGATGGTGCCGAGGCT
GCAAGGTTTATAAAAAAGATAAAGGAAATCATAGAGCAGCCAATGCTTTACATTGGTGATT
AA

**>*P. torridus* E3**

ATGGATTACGATGTCATTATCCTTGGGGCCGGTGCTGGTGGATATAAAGCTGCTGTTCATC
TTCTAAAAAATAAAAAAAGGGTATTAATGATAGAAAAGGAGAAATTCGGCGGTGAGTGCCT
TAACTATGGATGCATACCATCAAAGGCATTAATAGAGATGAGTGAGAGTATTTACTACCTG
AAGAACATGCCGGGCATCTCAATGGAATATAAAATAGATATGAAGGCATGGCAGGACTGGA
AGAATAGTATGGTTGCAAGGATAACTGGAAACGCAGAGAGGCAGTGCAAATCACTTGGTGC
TGATATATTATATGGCTTTGGAACATTAAAGGATAAAAACACCGTTAATGTCAATGGAAAG
GATTACACTGCGGAAAACATTATAATAGACACGGGATCATCGCCGGCAAGGATACCTGGAA
TAGACAATGTTTATTATAACAGGGAAATACTTGGCATTGATCATATACCGGAGAGCATTGC
AATCATAGGCGGCGGTTACATAGGCGTTGAGATAGGAACGGCAATGGCAAAGCTTGGTTCA
GATGTTTACATTATAGAGGCAAAGGAGAGAATACTTCCAGAGGTCAGCAGCGAGCTTTCAA
ATGCCGTTGATAAAAAATTAAGATCCATCGGTGTAAAAATAATGACATCCAGCAAGGTAAT
CTCTGTATCAAAAGATAAATATTACACTGTGAAAACAGAGAACGATGAGATAAAGTCTGAA
ATGGTTCTGATGTCTGTTGGAAGGATACCAAACACAAAAAATATAGGGCTTGAGAATCTTA
AAATTGAAATGGATGGAAGATTTATAAAGACCGATGAGCATAGAAGGACAAATGTAAAAAA
TATATACGCGATAGGCGATGTCACAACAGGACCGATGCTGGCACATAAGGCATTTTATGAT
GCATACATTGCATCAGAGAATATTCTCGGCAATGACGTAATAATAGACTATAAAGCAATGC
CATATGTGATATATACAGATCCTGAGATATCGTTCACAGGTAGTGTATCAAACAATTCAAA
GAAGATTATGACATCGGCAGTTCCAAGGGCACTTACAATGAATGAGGGCGATGGCTTTTTT
AGAATTTACTACAATGATGATGGAACAGTAACAGGAGCGGCCATCGCCGCGCCAAGATCAT
CAGAGATTATAACAGAGATAAGCCTTGCCGTTGAATCTTATTTAAATATAAATGATCTGCT
GCTTACGATACATCCACATCCAACAATAGCCGAGGGGATACATGATGACTTCGAAAATTAA

**> *P. torridus* LipB**

ATGATGACTTCGAAAATTAATTATTATATTGATCTTAAAAGATTTGATTACAAAAAATGCC
TTGACATACAGTACAGGCTTGTCTCAATGAGAAAGAACGATTTAATAGGTAATACAATAAT
TTTCGTTGAGCACGATCCTGTTTACACAATAGGCAGAAAGGCTGACCCAAGGAACTATTCA
AATGTAAATGTTATAAGAACAGACCGCGGTGGTGATATAACATATCACGGCCCCGGCCAGC
TGGTTTCTTACTTTATCTTTGATGTAAGAATAAATGGTAAAAAGGAGGTAAAAACCTTTCT
TAATAATATAGAAAACGTCTACATAAATCTACTAAAGAATCTTGGATACAGTGCATCAACA
GGCGATGAACCGGGCATCTGGATAGAAAAAAACGGTAAATTAAAAAAGGTTGCTTCCATAG
GCATGGCCATAGATGATTATGTATCATATCATGGTGTTGCACTTAACATAAACAAGGAGGT
TTTAAACGGATTTATAAAAATAAACCCATGCGGGCTCCCACCGGCCGTTATGGATTACATA
GATATTAAGAGAGATGATGCCATTGATTTAATAATAAGGGAGATTGAAAAAATATATGGTA
AATTCAATAAAATAGAATATGAGGGCCTGATCTCATGA

## Appendix 4: Publication